



Help Me #DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment

LU HE, Department of Informatics, University of California, Irvine, CA, USA

CHANGYANG HE, Hong Kong University of Science and Technology, Hong Kong, China

The wide spread of misinformation contributes to information consumers' excessed distrust of online information. To cope, information consumers are often actively involved in checking the credibility of information through self-researching or seeking help and opinions from experts and peers. While previous studies investigated the factors that affect people's perceptions of information credibility and how laypeople's judgements compare to experts, little is known about how the information credibility assessment work is performed and cooperated by individuals and communities in real-life, natural online environments. Through a qualitative study of an online community, r/DebunkThis, which is dedicated to information debunking, we found that online information debunking rarely followed a linear and straightforward path. Rather, community members, including the debunkers and the original posters, constantly negotiated, and interacted with each other to determine what to debunk and how to debunk. Individuals adopted various strategies to debunk information, such as questioning the credibility of the information source and citing authoritative external information. Community members supplemented with details and explanations, corrected others, requested clarifications, summarized high-level knowledge and skills, and interacted socially based on individuals' debunking explanations. Our study results broaden the understanding of debunking not only as an outcome but also as a learning and social process for community members to learn high-level debunking skills and form and enforce community rules. We provide implications for designing community and crowd-based information debunking systems which should recognize the complex, cooperative, and socially situated work of community and crowd debunkers. The design of such systems should therefore support not only labeling information as correct or not or simply sharing alternative information sources, but also community interactions and learning processes, as well as recognizing the labor of community debunkers.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing** → **Empirical studies in collaborative and social computing**

Additional Key Words and Phrases: Information credibility assessment, debunking, online community, social media

ACM Reference format:

Lu He, Changyang He. 2022. Help Me #DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment. Proc. ACM Hum.-Comput. Interact, 6, CSCW2, Article 413 (November 2022), 31 pages, <https://doi.org/10.1145/3555138>

¹ Authors' addresses: Lu He, lu.he@uci.edu, University of California, Irvine, Irvine, CA, USA; Changyang He, cheai@cse.ust.hk, Hong Kong University of Science and Technology, Hong Kong SAR, China.



This work is licensed under a Creative Commons Attribution International 4.0 License.

1 INTRODUCTION

While the ease of access to a huge amount of online information benefits users, the increasing prevalence of misinformation, those that are incorrect, distort facts, and deceive information consumers intentionally or unintentionally [17,34], is a growing concern. In the United States, information consumers view misinformation as a critical issue that causes confusion and threatens their trust in authoritative institutions and each other [4,40]. During the COVID-19 pandemic, the global internet was fraught with misinformation regarding origins of the virus, interventions, and treatments [23,56]. Misinformation has a concrete impact on people's beliefs, behaviors, and decision making [36]. In addition, marginalized populations and people with lower internet literacy are likely to be disproportionately affected by misinformation [45]. The spread of misinformation and low-quality information decreases information consumers' trust toward online information and increases their burden to fact-check information.

There is a rich stream of research that investigates various ways of fact-checking and debunking information and presenting corrected information to internet users [12,42]. One common strategy is to leverage the expertise and authorities of journalists and domain experts to identify and correct misinformation and explain it to the public [15,59]. However, manually checking the credibility of information is tedious and time-consuming for journalists and experts, and oftentimes fact-checks and debunking appear many hours later than misinformation [53]. In addition, the public may not always trust the explanation from journalists and domain experts, due to their affiliations and different political ideologies [7,8]. To tackle the growing amount of misinformation, another common strategy is to develop and deploy automated or semi-automated machine learning models to detect misinformation [1,37,61]. While many machine learning models have achieved satisfying performance in detecting misinformation, the increasingly diverse modality of misinformation [13,22] and the lack of transparency and explanation decreased the utility of machine learning models and internet users' trust toward them [52].

Besides journalists and experts who have intensive investigative experience and domain knowledge, laypeople are also found to be good judges of information quality [12,46]. Researchers and social media platforms study and implement crowd and community-based debunking systems that leverage and synthesize the wisdom of the crowd to identify and correct potential misinformation. For example, based on dissemination patterns of rumors and confirmed truths posted after the 2010 Chile earthquake, Mendoza et al. suggested that aggregating the crowds' reactions to questionable tweets could help identify rumors [62]. Zeng et al. found that Chinese internet users actively engaged in rumor verification on social media by leveraging various information sources such as foreign news and experts after a disastrous crisis and were better received by authoritative debunking [62]. Twitter started pilot testing of its community-based system named Birdwatch that enables users to add notes to posts that they deem as misleading [66,67]. The most popular social media platform in China, Weibo, has also solicited rumor reports from the crowd [68]. These efforts, however, haven't seen unexpected low-quality of debunking answers that lack evidence [64]. Therefore, it is imperative to study how to better design such community-based debunking systems.

While existing literature on understanding and improving laypeople's information debunking strategies and performances provides valuable insights [5,10,12,21,46], most regarded information debunking as an individual endeavor, without investigating the potential interaction among the individual debunkers, how it may affect the experience and outcomes of debunking online, and whether and how the debunking work is coordinated among community members. In addition, little

is known about how laypeople debunk information in a natural, online environment (e.g., online forum, Twitter, Facebook Page) where rich social interactions and social norms among the debunkers may play a significant role in shaping how information is debunked and the debunk is perceived by others. While information debunking shares similarities with other crowdsourced information tasks such as the Wikipedia project and collective sensemaking [25,29,39,44], it has a less established coordination structure and mostly relies on community-driven practices. Therefore, in-depth understanding of how the work of crowdsourced information debunking is coordinated online is required. Such insights will also shed light on how to design community-based debunking platforms to better support community debunkers' work.

To fill the gaps, we conducted an exploratory, in-depth qualitative analysis of an online community that is dedicated for users to seek and provide information credibility assessment, or information debunking. We found that individuals adopted various strategies to identify and explain whether the information posts are credible. In addition, community members actively supplemented with details, requested clarifications, corrected others, summarized high-level information, and socially interacted in the debunking process. Our work complements previous studies on individual layperson's information debunking work by providing a nuanced and detailed description of how individuals and the community perform information credibility work, both individual and collaborative, in a natural, online setting. We highlight that, such work is rarely linear and individual, but rather complex, collaborative, and social. Our work also broadens our understanding of debunking not only as a goal with binary (right or wrong) outcomes, but instead as a process through which individuals as community members (among debunkers and between debunkers and original posters) collaboratively assess information credibility, learn about debunking skills, initiate social interactions, and form and reinforce community norms. We suggest that community and crowd-based debunking systems should support not only the basic functions of labeling information as correct or not, but also collaborative, social, and learning functions. In addition, community-based debunking systems should recognize the complex work behind the input from community debunkers, instead of simply treating them as free labor to label information credibility.

Our work contributes to the HCI and CSCW community by 1. providing a nuanced and detailed description of the individual and collaborative work carried out in community-based debunking; 2. broadening the concept of information debunking beyond a goal but as a process that initiates and supports collaboration, social, and learning functions; and 3. offering design implications into designing community-based information debunking systems.

2 RELATED WORK

2.1 Journalist and expert-based, and automated information credibility assessment

The growing amount of misinformation has contributed to internet users' pervasive distrust toward online information (e.g., news, social media) [35,60] and given rise to the need to check information credibility before making decisions based upon the information [14]. The pressing concern of misinformation has led to numerous widely used fact-checking platforms run by non-profit-organizations, journalists, or commercial entities, such as FactCheck.org [61], Snopes [62], and Tencent COVID Fact-check [63]. Major technology companies and social media platforms have also deployed fact-checking services of their own or with third-party collaborators to flag online news and articles as fact-checked or not, such as Google [55] and Facebook [64]. Another strategy is to leverage the expertise and authority of journalists and domain experts to identify and correct misinformation and explain it to the public [15,48]. Journalists investigate the information in-depth,

solicit expertise and experiences from multiple credible sources, and produce fact-checking reports. However, manually researching and checking the credibility of massive online information is tedious and time-consuming for journalists and experts, and oftentimes only selected information was debunked. In addition, the public may not always trust the explanation from journalists and domain experts, due to their affiliations and different political ideologies [7,8].

To tackle the growing amount of misinformation at-scale and in-time, another common strategy is to develop and deploy automated or semi-automated machine learning models to detect misinformation. The advancement of more powerful models and the incorporation of behavioral theories have improved the performance of machine learning models in detecting misinformation in various fields. For example, Zhao et al. developed a machine learning model using behavioral, topical, and emotional features, and successfully detected 85% of the health misinformation from an online community [54]. Transfer learning also enabled machine learning models to adapt to different domains. For example, Bojjireddy et al. developed machine learning models that were trained on three large public datasets before COVID-19. The models reached promising performance on COVID-19 datasets, which demonstrates their adaptability to different domains [6]. While machine learning models are promising to identify misinformation at scale, they also face several challenges. First, many machine learning models are developed on textual data (e.g., social media posts, news articles), yet misinformation has increasingly diversified modalities such as images [13] and videos [21,29]. Second, most machine learning models lack transparency regarding how the decision of whether the information is incorrect was made, which decreases users' trust toward model results [41].

Despite the value of journalist and expert-based information debunking and automated machine learning debunking, the challenges of scaling up expert debunking explanations and lacking transparencies and explanation of machine learning models preclude their wide utility to identify and correct online misinformation efficiently. An alternative is to leverage the help from laypeople, or the crowd and the community to debunk information.

2.2 Wisdom of the crowd: Community-based information crowdsourcing and debunking

2.2.1 Community-based information crowdsourcing.

With the internet being increasingly democratized, people obtain information increasingly from their peers, instead of only from experts. People who are interested in sharing information and knowledge with a wider audience form online communities or organizations such as Wikipedia and subreddits. Crowdsourcing information and generating knowledge as a community has received extensive and sustained attention from the HCI and CSCW community. Below we review three strands of literature that are most relevant to community-based information debunking: collective knowledge development, collective opinion evaluation and persuasion, and collective sensemaking. One of the most prominent examples of crowdsourcing information tasks is collective knowledge development. The most popular example, Wikipedia, hosts information about a wide variety of topics that is contributed, curated, and maintained collaboratively by users that are only connected through the internet. While information has been commonly regarded to be objective, the creation, curation, and maintenance of information in collaborative online communities by the crowd have been shown to be subject to various social and community norms, social processes, and biases. This is in part because the work involved in crowdsourcing information needs to be distributed, spontaneously or planned, among many users with different backgrounds and ideologies. For example, Menking and Erickson noted the gender gap in Wikipedia contribution where women users carried out significant emotion work to participate [39]. Successful and efficient

crowdsourcing information requires coordinating the work among information contributors, which has been extensively studied in the HCI and CSCW community. For instance, Kittur and Kraut found that simply adding more users to contribute to Wikipedia articles did not always improve the information quality but often needed to depend on the coordination method used. Their findings suggest that the power of crowdsourcing information lies in not only many contributors but also effective and appropriate coordination strategies [25].

Besides Wikipedia, there are many other emerging online platforms and communities that gather crowdsourced information from a wide range of users. Collective opinion evaluation and persuasion, through which community members assess others' beliefs, derive reasonings and try to persuade others, becomes a common scenario. For example, on r/ChangeMyView, users post their viewpoints related to heated topics such as social issues and solicit arguments from other users to challenge their viewpoints [24,43,55]. Jhaver et al. studied what motivates people to contribute to the forum and post arguments to challenge others in a civilized manner, what maintained a civil environment, and whether the crowdsourced arguments are effective in changing people's minds [24]. Priniski and Horne found that on ChangeMyView, users are more likely to change their views when discussing non-sociomoral issues and when they are presented with more evidence [48]. Musi et al. found that concession in arguments does not lead to the change of views [43]. Besides investigating factors that contribute to successful persuasion, Srinivasan et al. studied whether and how content removal can effectively make users being more civilized in the ChangeMyView forum [55].

Another extensively studied crowdsourced information activity is collective sensemaking. During times of crises when credible information is highly valuable yet scarce and constantly changing, people rely on second-hand information, speculations, and even rumors to collectively make sense of the situation and make decisions. For example, during the Zika outbreak, users on Reddit formed and discussed conspiracies theories to collectively make sense of the crisis [29]. Patients and caregivers also exchange information with each other on online health communities to collectively make sense of health conditions, coping strategies, and patient journeys [38].

In summary, community-based information activities such as Wikipedia and collective sensemaking rely on complex sociotechnical systems and operate upon social norms and processes that require effective and carefully designed coordination strategies to make the crowdsourcing tasks work [19,44].

2.2.2 *Community-based information debunking.*

Besides journalists and experts, laypeople are also found to be good judges of the quality and validity of online information [12,46]. Extensive research has investigated the factors that contributed to individual users' perceived credibility of online information (e.g., Morris et al.[41], Lazar et al. [31]), and how such perceptions differ in different cultural contexts [52]. Studies have also analyzed how laypeople's performance in information credibility assessment differs from experts and suggested assigning tasks that are tailored for crowd debunkers with different backgrounds [5]. Information credibility assessment from the crowd can also be integrated into social media post ranking to inform information consumers which posts are more trustworthy. Major social media platforms such as Twitter have started pilot testing of its community-based debunking system, Birdwatch, which allows users to identify misleading posts and submit notes to explain their reasoning [66,67].

While these studies provide rich insights into how users evaluate information credibility, they only shed light on individual users' behaviors and perceptions. As users are increasingly connected

with each other with shared interests and goals, they often collaboratively approach such tasks because the community (or the crowd) often has members that possess more expertise, knowledge, and experiences in the subject matter. For instance, Zeng et al. reported that after the Tianjin explosion, Weibo users collaboratively debunked rumors by utilizing external sources of information from foreign websites and experts [62]. However, the process through which community members collaboratively debunk information, especially in a natural online environment where rich social interactions and norms are likely to be influential, is less explored.

2.2.3 Differences between crowdsourced information debunking and other crowdsourced information tasks.

Though online debunking bears many similarities with other crowdsourced information tasks, there are important differences that may lead to different community dynamics, coordination practices, and requirements for platform support. First, knowledge-production work such as Wikipedia has different goals from debunking. As Wikipedia states, its goal is to create “a widely accessible and free encyclopedia” [65]. It is therefore expected and organized to produce comprehensive knowledge and new knowledge will be added constantly to the existing body. In contrast, debunking does not aim to be comprehensive but rather deals with what emerges. The goal for debunking, instead, is to evaluate claims and information, assess its credibility, and correct if possible. Therefore, though both Wikipedia and debunking solicit input from the community, they have inherently different goals, which may in turn lead to distinct communication patterns and coordination strategies. As briefly reviewed in Section 2.2.1, there is extensive literature on the communication structures and patterns in knowledge-production for Wikipedia. For example, Schneider et al. discovered communication patterns on Wikipedia such as coordination, guidelines, and editor notes [51]. In contrast, less and little is known about such communication patterns in community-based debunking. Most debunking literature still focuses on the quality of debunking outcomes, without exploring the communication processes behind them. This lack of investigation may be in part due to the fact that many information debunking platforms such as Twitter’s BirdWatch are very recent. Even though studies about how to effectively debunk information and persuade the potentially misguided public have been myriad, most are conducted in lab environments without investigating how information debunking is carried out and perceived in real, natural online environments where social norms and processes can be significant.

Wikipedia also has a dedicated platform with specific interfaces and functions that support the collaborative work behind it. For example, the Talk page facilitates editors’ work to change the articles [51]. The produced knowledge is displayed in a Wikipedia article with organized sources and edit histories. In contrast, most existing community-based debunking still relies on general-purpose platforms such as Reddit. Even though Twitter has recently launched a dedicated functionality BirdWatch, it is still primitive and demonstrates deficiencies in supporting community-produced high-quality results. It is unclear how community debunkers invent and adopt strategies to coordinate their collaborative work in debunking on these general platforms.

Online information debunking is also different from collective sensemaking which is often community-based and involves crowdsourcing information from a wide range of users. For example, in collective sensemaking, there is generally no pre-defined questions or clear goals, while information debunking often starts from a narrow, focused question such as whether a statement is true or not. In collective sensemaking, community members may simply want to make sense of an ongoing crisis. Therefore, even though community members both contribute their knowledge to the process of collective sensemaking and information debunking, the goal and dynamics may be

distinct in these two crowdsourcing information tasks. Such differences are less explored in existing research.

In summary, even though several other crowdsourcing information tasks such as the Wikipedia project and collective sensemaking have been studied in the HCI and CSCW community, the different goals and platform support may bring unique challenges for community-based debunking. The insights from extensive Wikipedia and collective sensemaking studies may therefore not fit the needs of community debunkers. How community debunkers on general platforms coordinate their work and how such platforms should be better designed need further studies.

2.3 How online platform design shapes and constrains online information tasks

Users with shared identities, interests, and goals gather to form online communities or interact with each other on social media platforms to exchange informational and emotional support. Online communities have been extensively studied in areas such as health (e.g., women's health [30], public health crisis [18], and cancer [33]), games [27,28], education [16], creative critiques [26], and political issues [44]. Besides online communities that are more coherently organized and with established norms and dedicated moderators, social media also affords collective activities among the crowd such as collective sense-making during times of crises [20], sharing and learning for data scientists [54], and collective actions and social movements [57,58].

This rich stream of literature in the HCI and CSCW community investigates how online community and social media users exchange information and support and organize collective activities. One relevant strand of literature studies how the design and affordances of online platforms support, affect, and even preclude users' activities. For example, Leavitt and Robinson reported that Reddit users used the functions on the platform such as voting to appropriate the visibility of information during crises, where in-time and accurate information is crucial [32]. Liang found that the structure of Reddit threads (e.g., thread depths, number of nested comments) also affects the ratings of knowledge sharing threads [35]. Besides the explicit functions, platform policies such as moderation is also important in understanding how users interact with each other and contribute to their online communities. For example, Gilbert found that in r/AskHistorians, visible moderation is often viewed as censorship and has created challenges for volunteer moderators [14]. In turn, the norms and cultures of online communities may also affect how the functions are used by users and the construction of the communities [47].

With this abundance of literature on the dynamics between online platform design and affordances and user interactions, however, little is known about the cases in crowdsourced information debunking in online communities. Even though previous literature termed debunking on social media and online communities as collective activities, most approached them as disjoint, individual work [10,63]. It is unclear how the debunking work is cooperated among community members, how they interact with each other in the debunking process, and how the interactions and social norms on online communities and social media platforms may affect online debunking, and how user interactions and platform affordances are interrelated.

3 MATERIAL AND METHODS

The objective of our study was to understand and unpack the individual and collaborative work among community members when debunking information online in a natural setting. Therefore, instead of distributing surveys or conducting lab-based experiments, we looked for online

communities and social media platforms where rich discussions around information debunking happen.

3.1 Data source and collection

3.1.1 Data source.

To locate platforms that host active discussions of information debunking, we searched multiple popularly used social media platforms including Twitter, Facebook, and Reddit, using keywords “debunk”, “information credibility”, and “misinformation” in January 2021. We found that Twitter and Facebook (public pages or groups) include mostly news or posts that claim rather than discuss which pieces of information are not credible. We did not find enough posts that included debunking discussions which provided detailed reasoning. Finally, we located a subreddit called r/DebunkThis on Reddit, which is dedicated to information debunking. The subreddit was created in 2010, and as of June 2021, it has 20.1k members. The description of the subreddit illustrates its mission: “We are an evidence-based subreddit dedicated to taking an objective look at questionable theories, dodgy news sources, bold-faced claims, and suspicious studies.” As of June 2021, the subreddit has eight moderators. The community also instructed how users should post to seek debunking: “ALL post titles use the following format: “Debunk This: [main claim here]” Posts Must include (in the body of your post): • Between 1 and 3 specific claims that you want debunked. • Your post must include at least one source. • For video links, time stamps to specific parts that you want debunked. * Topic must be at least 2 months old. This information will give other users a specific point to look into and will increase the chances people replying to your post.” Moderators reserve the right to remove low-effort posts.

3.1.2 Data collection and descriptive summaries.

We used the Python package PRAW to crawl all posts and associated comments on r/DebunkThis on March 15th, 2021 [69]. The package is built on the Pushshift application program interface (API) and has been used for collecting data from Reddit in many other studies [9,30]. All posts in the subreddit were collected. The data included title, textual body of the original posts and all associated replies, time stamps for posts and replies, and usernames of the original posters (OPs) and users who replied. Posts and replies that were deleted by their authors, removed by moderators, or archived were also collected but excluded from further analysis.

In total, 4,647 threads and 44,545 associated replies were collected after those deleted, archived, or removed were excluded. There were 2,531 unique OPs who sought debunking and 4,241 unique users who replied. On average, each post received 9.56 replies from the community. The data span from May 14th, 2010, when the subreddit was created, to March 15th, 2021, when we started our data collection. The community has witnessed a sharp increase in the number of posts since the start of COVID-19, with more than 200 posts for each month from April 2020 to June 2020.

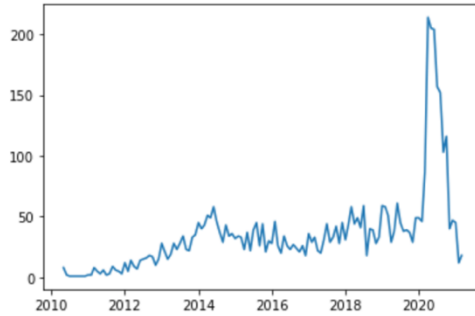


Fig. 1. Temporal trend of the number of threads in the subreddit.

3.2 Data analysis

To understand the strategies and work taken on by individuals and the community to debunk information, we adopted a grounded theory approach [11]. Two authors independently coded 40 random threads to generate initial codes. For each thread, they first read the title and description of the debunking request and went through the content of external source(s) (e.g., blogs, news articles, videos) that the users wanted to debunk if any links were given. Such a procedure provided a basic understanding of the context of debunking requests. Then, they read through the replies, focusing on how users identified the credibility of the information, explained their reasonings to others, and how other community members contributed to this process. Two authors generated initial codes on how users raised the debunking questions, how debunkers individually debunked the questions with convincing explanations, and how different debunkers in the community interacted with each other to debunk the information. They then discussed and compared their initial codes to reach a consensus and produce a codebook. The two authors then went back to the data and coded more threads, applied the codes to new threads, and recorded other emerging codes. The coding was finished when no new codes emerged. After this iterative coding process, two authors coded 108 unique threads, containing 2,736 replies contributed by 464 unique users. To verify whether COVID-19 related threads are overrepresented in our study sample, we additionally coded the topic of the thread. We found that while 37.0% (40 out of 108) of the threads are posted after 2020, only 9.3% (10 out of 108) of them are related to COVID-19. Therefore, we conclude that COVID-19 does not affect our study sample. Generally, society (N=26), occult (N=23), health (N=22), and politics (N=12) are the four dominant topics in our study sample. The remaining threads cover different subareas of science, such as physics (N=6), ecology (N=6), and biology (N=4).

To verify whether COVID-19 related threads are overrepresented in our study sample, we additionally coded the topic of the thread. We found that while 37.0% (40 out of 108) of the threads are posted after 2020, only 9.3% (10 out of 108) of them are related to COVID-19. Therefore, we conclude that COVID-19 does not affect our study sample. Generally, society (N=26), occult (N=23), health (N=22), and politics (N=12) are the four dominant topics in our study sample. The remaining threads cover different subareas of science, such as physics (N=6), ecology (N=6), and biology (N=4).

3.3 Ethical considerations

We adopted several measures to protect the privacy and safety of the online community we studied. First, we determined that our study was qualified for self-exempt based on the self-exempt tool provided by our university Internal Review Board (IRB). Our study falls under the self-exempt category of secondary analysis of publicly available data. The IRB self-exempt tool instructed those studies qualified for self-exempt are considered as confirmed by university IRB and no review is required and will not be provided. We also did not interact with any human subjects in this study in any ways (e.g., replying to posts, adding community members as friends, chatting with community members). In addition, the data collected were stored in computers protected by passwords and only accessible to the research team. The data are not shared with people outside of the research team nor publicized online. With these safety measures, we believe that we introduced little to minimal risk to the online community we studied.

We acknowledge that complying with IRB instructions is not always the best practice for ensuring the safety of online communities studied. Based on findings and suggested practices from literature on the ethics and safety of conducting research on Reddit [49], we did not use direct quotes but instead paraphrased all posts that we quoted in the paper so that they cannot be searched and traced back to the original posts. We also did not directly use any real usernames or user IDs in this paper.

4 RESULTS

We found that rather than a linear, straightforward, and individual process, debunking information in the online community is a complex, iterative, and social process that is cooperated and negotiated among community members. Individual community members and the community collectively adopted various strategies to identify what to debunk and how to debunk. First, they decided and negotiated the debunkability of the information presented. Second, individuals used various strategies to both debunk the information based on its own characteristics and based on external sources or personal knowledge. Third, community members collaboratively debunked through supplementation, correction, request for more information, summarization, and social interactions around individuals' debunking explanations. We describe the individual and community work carried out in information debunking in the following sections. We use **OP** (Original Poster) to refer to users who posted to seek credibility assessment from others, and **D** (Debunker) to refer to users who provided debunking explanations to the OPs and the community. Finally, we used a sample case to illustrate how various individual and collaborative community work collectively play out to debunk information, which involves more than simply judging a post is credible or not.

4.1 Deciding and negotiating debunkability

Before debunkers delved into debunking the information posted, they often first decided whether the information was clear and suitable to be debunked, which we term as debunkability. When posting in the subreddit to seek information credibility assessment from the community, many OPs simply posted a link to a website, or a piece of text copied from other sources, even though the community rules clearly forbid such posts. In our data, we observed that the work of negotiating what to debunk and whether the information is debunkable was commonly carried out by the community members before any in-depth debunking explanation was provided. For example, OP1 asked the community to help debunk a blog that stated, "some vegetable leftovers should not be reheated otherwise they will poison people". D1 replied that "There were no citations, no scientific credentials, nothing to refute.." D2 agreed and replied "Yes! Pretty much. Nothing to actually debunk

as nothing was presented..." The debunkers considered the information presented here not debunkable because it did not provide any clear statement to be debunked.

Besides posts that lacked clear statements, those that involved ideologies and moral judgements were also deemed not debunkable by most debunkers. For instance, D3 explained that "If it's opinions of what people believe to be morally wrong, you can't really debunk it." In addition, while some posts are deemed as undebunkable by all or most debunkers, in some cases, the determination of debunkability also varies across the debunkers. When labeling some posts as not debunkable, debunkers interacted and negotiated with the OPs to ask for clarification to their posts to make them debunkable. For example, D4 asked P2 "What is there to debunk?" and OP2 specified that "I was hoping for someone to debunk the technology itself with a Samsung smart tv and the government motive specifically. I wasn't sure about that." With the clarification from OP2, D4 and other debunkers were able to provide detailed debunking explanations. In some other cases, negotiating debunkability can be more complex and iterative, with several rounds of questions and clarifications.

4.2 Individual work of information credibility assessment

After deciding a post is debunkable, individuals set out to provide detailed explanations to debunk the posts through various strategies, including debunking based on specific characteristics of the information such as the credibility of information source and logic, debunking with the support of external information, debunking with personal knowledge, and combining multiple strategies.

4.2.1 Debunking based on specific characteristics of the information.

Debunking directly based on the characteristics of the information is a typical strategy in assessing information credibility. Without fact-checking whether the information is correct or supported by scientific evidence, some debunkers arrived at their conclusions by screening surface characteristics of the information. One characteristic is whether the information is from a credible source. Debunkers considered multiple factors for deciding the credibility of information sources, e.g., whether the author has relevant professional training and background, whether the author has unbiased and neutral viewpoints and affiliations, and whether the author has had a history of creating and disseminating suspicious information. For instance, D5 identified an article authored by a researcher which claimed, "climate change is not real" as not credible. D5 explained that the researcher was not a credible source because he has affiliations and funding sources which may affect his neutrality as an information source:

"The author clearly has an agenda and is funded by companies to push that agenda. I won't take this as a credible information source.."

Information that was not accompanied by relevant supporting material, or only by suspicious and intentionally manipulated reference was also a signal for debunkers. The quote below exemplified when a debunker found the information not credible because of the lack of credible evidence:

"Citing sources that don't relate to the place they're cited is a major dishonesty. The fact that this article does so repeatedly is a huge red flag."

Debunkers also found information suspicious if details of the content seem to be edited intentionally or cherry-picked. For example, a poster sought debunking for a video that was claimed to capture a supernatural existence. A debunker found the video not credible because:

“The video looks like it’s been edited and slowed down. If you look closely, there’s a weird dark mark that doesn’t look natural in an unedited video.”

Debunkers may spot basic logic flaws of the information that led to their questioning of the information’s credibility. Logic flaws such as exaggerated statements, conflicting statements, unsubstantiated inferences, and drawing conclusions that are too absolute without considering alternative hypotheses are considered as “red flags” for debunkers. For example, a debunker found a post which stated “*Company X is engaged in sex trafficking trade and selling children*” to be suspicious because of logic flaws and many other possible explanations unstated:

“The post just contained unrelated and random facts and put them together using suspicion and paranoia to come to a wild and unsubstantiated conclusion. There are many other possible explanations in fact, which the post clearly avoided.”

Without searching for external information or leveraging domain and personal knowledge, debunkers sought to assess the credibility of information based on surface characteristics of the information presented. They relied on these characteristics that serve as “red flags” to determine whether the information is suspicious, even though the information itself was not actually debunked.

4.2.2 *Debunking with the support of external information.*

Debunkers commonly cited external information that they deemed credible and of high-quality, such as peer-reviewed scientific papers, fact-checking platforms, mainstream news websites, Wikipedia, blogs, or even posts from other subreddits to debunk and support their arguments. Some debunkers used external information to directly contradict the claims made in the original information. For example, a debunker cited a link to a medical journal supporting an argument to debunk the claim that abortion kills more women than childbirth:

“The risk of death associated with childbirth is approximately 14 times higher than that with abortion. Similarly, the overall morbidity associated with childbirth exceeds that with abortion. source: URL”

Some information has been debunked elsewhere (e.g., another subreddit, fact-checking platforms), and debunkers often cited these external sources in their posts. For example, a debunker directly posted a link to a fact-checking platform that debunked the claim that Dr. Fauci wrote that hydroxychloroquine was effective for coronavirus in 2005:

“It has already been debunked many times elsewhere. Here’s one: <https://www.politifact.com/factchecks/2020/may/06/blog-posting/dont-fall-conspiracy-about-dr-anthony-fauci-hydrox/>”

While debunking with external information sources is often short, it was well-received by others because debunkers usually chose sources that are highly credible and authoritative.

4.2.3 *Debunking with personal knowledge.*

Besides external information, some debunkers may also directly use their personal knowledge to debunk. Specifically, we identified two types of personal knowledge that debunkers can leverage: domain knowledge and situated knowledge. Domain knowledge is usually from a specific, specialized discipline or field where a debunker works or is trained in such as physics and medicine. Debunkers commonly leveraged domain knowledge when they would like to unpack the

mechanisms of the information and identify incorrect claims. For example, a debunker raised doubts on a non-peer-reviewed paper regarding "free energy":

"Based on my understanding, Zero Point Energy represents the lowest point at which matter can reach. In electronics terms it's ground. It's the reference point closest to 0, so it sounds to me like someone thinks you can get something from nothing"

Situated knowledge is gained because a debunker, or friends and families of the debunker has personally experienced specific situations and thus obtained relevant knowledge. For example, a debunker refuted the claim that 911 was fake and fabricated by the government:

"My mother saw the plane hit the tower, in person. She was working nearby. She went up on the roof of the building to see the fire/smoke. How can this be fake?"

Similarly, another debunker testified that the death rate of COVID-19 presented in a chart was accurate by providing his/her personal experience:

"The numbers are very close. I know this because I work as a nurse in ICU."

Even though utilizing personal knowledge to debunk is relatively easy for debunkers, some OPs have expressed their concerns about this strategy. For example, when a debunker leveraged personal knowledge to evaluate a study as "a test with a small sample size", the OP questioned, "Is almost 100 people a small sample? (no sarcasm intended)".

4.2.3 Debunking with multiple strategies.

When there are cases where the information has multiple aspects that are suspicious, the same debunker may leverage more than one of the previously mentioned strategies and combine them to debunk the information. For instance, an OP sought debunking to the fatherless effect on crime rates and social issues. A debunker leveraged multiple strategies we identified before to question this claim:

*"If the questions are : 'Is it true that children from fatherless households are significantly more likely to 1) be involved in criminal behavior, and 2) to have much lower educational attainment?' Then may I suggest checking the following [r/asksocialscience](#) thread.
... Family structure may be what we notice first, but it often leads to ignoring other important factors such as family history and family dynamics. (And, broadly speaking, context.) It is also often associated with the incorrect assumption that all 'single-parent families' or 'fatherless families' are of the same type."*

This debunker first provided an external information source (a post in another subreddit) that had already discussed and analyzed the relevant issue. The debunker further leveraged personal knowledge to indicate the complex influencing mechanisms of family context to crimes, instead of only considering the fatherless factor.

In another case, an OP asked for debunking to a post shared to her through WhatsApp, which stated that COVID-19 was made and spread from a lab. A debunker opposed the claim by providing a debunking explanation from an external fact-checking platform, questioning the credibility and motivations of the information source, as well as leveraging domain knowledge to explain why the virus was unlikely to be lab-made:

“First, take a look at this: <https://www.businessinsider.com/coronavirus-lab-manmade-myth-debunked-2020-6> The details of the virus’ genome make it extremely unlikely to have a laboratory origin. Then there’s the lack of that person’s credibility, given that she only published scientific papers on a different subject. Also, the virus’ genome makes it extremely unlikely to have a laboratory origin.”

4.3 Community work of information credibility assessment

Besides individually contributing to information debunking on the subreddit by leveraging the previously mentioned strategies, community members also collaboratively improved information debunking and socially interacted with each other. We distinguish community information credibility assessment work from its individual counterpart based on whether the debunking is related to or responding to another user’s post. We note that not all replies to the original posts or other replies were regarded as community work. Some debunkers directly replied to the original posts or other debunkers, without any references to specific aspects of other debunkers’ posts. Instead, we considered community work in replies that referenced other debunkers’ points and initiated rich discussions and interactions which collaboratively contributed to debunking the information or improving the quality of existing debunks.

Among the 108 threads that we coded, 80 (74%) threads involved multiple community members in collaboratively assessing information credibility or initiating social interactions during debunking. The prevalence of collaboration and social interaction among community members indicates that information debunking online is not only an individual endeavor but also a collaborative process of a community that is governed by negotiated community norms and initiated and fostered dynamic social interactions. In this section, we describe how community members collaboratively supplemented, requested, corrected, debated, summarized, and interacted socially around individuals’ debunking posts.

4.3.1 Supplementation.

If a debunker mostly agreed with posts by other debunkers but believed the reply was not adequately comprehensive, rigorous, well-demonstrated, or had other points that were worth mentioning, the debunker might improve it with supplementation. We note that supplementation is different from a debunker simply replying to the OP with his/her own explanation that is different from other debunkers’. Supplementation is different because it is grounded in acknowledgement of other debunkers’ explanations and directly built upon it, instead of starting one’s own explanation. The connection and continuity with other debunkers’ explanations are what distinguishes supplementation from disjoint, individual explanations.

For instance, a debunker refuted a chart that claimed white households in the US earned less than many Asian households by pointing out that there was selection bias, i.e., only Asians who successfully immigrated to the US are included in the chart and they are most likely to be highly skilled workers and thus earn more. Another debunker, while agreeing with what the debunker explained, supplemented by pointing out another flaw of the chart that challenges its validity:

“You are right. In addition, they also represented specific Asian ethnicities such as Indian and Chinese separately but have lumped all White ethnicities together as a single data point. Another red flag.”

In another case, a debunker shared external information sources to refute the claim that the government knows that cannabis cures cancer yet intentionally hides the fact. Another debunker supplemented the answer by providing domain knowledge on cancer:

“Also, the claim of curing cancer is a huge red flag. There are many varying types of cancer and thousand different ways of treating them. It is impossible to just have one cure for cancer in general.”

Debunkers supplemented more than just additional details or perspectives to other debunkers' explanations to strengthen them. They also supplemented high-level reflections or even warnings of debunking methodologies in some cases, to make the debunking more rigorous by pointing out its limitations that readers should be aware of. For example, a debunker sought to debunk a claim by pointing out the author of the claim is a racist and thus the statement is biased. Another debunker, while acknowledging the strategy of discrediting information authors was inevitable, also supplemented with its limitation:

“Just because a racist or someone not so credible or unbiased said it, doesn't necessarily mean their data and statistics are wrong.”

Oftentimes, supplementation of existing debunking explanations received agreement and recognition from the debunker whose post was supplemented. They either expressed gratitude (e.g., *“Thanks for adding to it. Didn't think about it that way but I agree.”*) or directly edited their previous explanation.

4.3.2 Request.

Debunking a claim often requires a clear definition and common grounds of statements, terminologies, and contexts in the information, especially when the OPs only posted a link, a screenshot from social media platforms, or a vague statement made in a blog. Therefore, it was common for the original posters and debunkers, or among the debunkers, to request definitions of terms, clarifications, and further explanations. For example, an OP sought debunks to a video that shows an abortion process. A debunker requested more information to decide whether the video is valid:

“Could you let me know the context? Circumstance of abortion? Trimester?”

Sometimes the OP was not satisfied with debunkers' explanations and requested further debunking. For instance, a debunker opposed the report posted by the OP which stated that climate change is not real. The OP was not convinced by the explanation and requested further debunking:

“But it didn't really refute the statements in the article. I think the interpretation of facts in the article, or the facts themselves, could be refuted in some way?”

Other debunkers may also request more information from a debunker if they found the debunker's claim to be unsubstantiated. For example, a debunker found the chart presented by the original poster to be true. Another debunker requested justifications for this belief:

“Is it only based on your opinion, or on some factual information? Do you have any sources to support it?”

Besides debunkers requesting clarification from OPs or evidence from other debunkers, OPs may also request from debunkers. For example, when OPs are not satisfied with the debunking answers,

they may request additional debunks of specific methods. In the following thread, the OP was not convinced by the debunker's answer that the author was not credible. Rather, the OP requested debunk that directly refutes the data presented in the original information:

"I am looking for actual debunk of the trend not the authors have an agenda etc, a agenda doesnt stop data being true or false. I need evidence of the data being false."

4.3.3 Correction.

Debunkers may also spot mistakes or deficiencies in other debunkers' explanations, and they often pointed out the issues and provided corrected answers in order not to mislead others. Such correction either aims to improve the debunking quality, when debunkers hold similar opinions with the existing explanations, yet find them not rigorous; or tries to "debunk the debunking", when debunkers believe the current argument totally fallacious.

Some debunkers corrected others by pointing out the external information sources they referenced are not trustworthy or neutral or that they misinterpreted certain aspects of the information. For example, a debunker found that another debunker misinterpreted the statistics in the original post and thought the numbers were estimated. The debunker therefore corrected the mistake:

"It's not an estimation but actually real data out of a study population of 260,000 inmates in the US."

Some debunkers identified the non-rigorous statement and provided corresponding evidence. In the following instance, a debunker claimed that "Autism is a rare and poorly-understood disorder", and another debunker corrected this claim by sharing statistics from authoritative sources,

"For your information, in the U.S., autism spectrum disorders are a lot more common that you might believe. The CDC has a statistic that 1 out 68 children have autism. The Autism Society estimates the worldwide rates at about 1% of the population. It's indeed poorly understood, as you say, because there are so many types that vary in their severity and symptoms."

Besides correcting factual details as the examples above demonstrated, debunkers also corrected the reasoning behind other debunkers' replies, when they believed that the debunker arrived at incorrect conclusions even with the right data. For example, a debunker used statistics of mortalities from Sweden to infer possible mortalities worldwide. Another debunker, while acknowledging that the statistics were valid, corrected the debunker that it was not a correct inference because the testing coverage and lockdown policies differed substantially between Sweden and other countries:

"This is a great explanation, but unfortunately you made the same mistake, that is, misusing factually correct numbers. Unfortunately, we cannot simply compare mortalities between Sweden to approximate the effectiveness of lock downs and infer mortality rates. I'm simply pointing out that you cannot draw your conclusions from the data you are using, even though the data is valid."

Generally, the community-based correction involved a crowdsourced reviewing process, which helped to reduce the biases or mistakes of individual debunking.

4.3.4 Debate.

It was a common scenario that different debunkers held opposite opinions towards the raised problem for debunking, especially when they had varying levels of domain knowledge and were possibly influenced by preconceived biases. Such disagreements would raise debates, when two debunkers proposed personal judgment, provided corresponding evidence, and uncovered the

other's flaws (if any) in the nested replies. Here is an example when two debunkers had conflicts during debunking a peer-reviewed article about sasquatch footprints:

Debunker 1: Not even one fossil? Not one to show that another upright walker lived on the western hemisphere before humans crossed the land bridge...

Debunker 2: What fossils have been unearthed in the Pacific Northwest? I mean of any animals? That's not a reason to blow off a pervasive phenomenon with an overwhelming abundance of physical and anecdotal evidence.

Debunker 1: Uh... the vast majority of dinosaur bones have been found in North America. Including Northwest but especially in the rocky mountains. But alas, no prehistoric humans. Totally different era and evolutionary chain. [LINK 1](#). [LINK 2](#)."

The debate can deepen the debunking discussion with more comprehensive perspectives and more abundant evidence. However, when preconceived biases existed, debunkers might fail to convince each other. Also, in some cases, debates might evolve into off-topic quarrels. To this end, other community members played a significant role in crowdsourced evaluation and moderation. For example, in the above example about sasquatch footprints, some other debunkers further replied to support debunker 1, e.g., "footprints are easily forged and doesn't qualify as a specimen... [LINK](#)", which helped to reach a consensus in the thread. In another case when debates developed to abusive language, community members reminded the community norms, "NO hominem attacks".

4.3.5 Summarization.

For threads that contain multiple debunkers' explanations that may be challenging for other readers to immediately grasp the key points, some debunkers, or even the OPs would post summarizations of these explanations to help others comprehend. When some debunkers only posted a link without many explanations, others may read through content of the website that the link directed to, summarize, and share the key take-aways with others. For example, in a thread where the OP sought debunk to the claim that 5G is a health threat to people, a debunker posted an external link with no explanation. The OP was convinced by the content of the external information source and summarized the key take-aways for others:

"Thanks! Looks very credible. Major takeaway points are: 1. The radio wave emitted by 5G is non-ionizing; 2. the only non-ionizing radiation which could be damaging is microwave, but that's been deliberately altered to do that exact task in controlled devices; 3. It is indeed listed by the WHO as a possible carcinogen, but on the same list as caffeine."

Such summarized information was valued by OPs and other community members, and sometimes even got shared outside the debunking community. For instance, in a thread trying to debunk some racism-related data, a debunker concluded different flaws detected by other debunkers. The OP replied, "Thank you! I caught it on Facebook. Can I use your answer to refute my friend?" and got the debunker's consent.

Sometimes, debunkers also summarized high-level debunking skills, i.e., meta-knowledge of debunking, that they believe could help others identify similar misinformation as presented in the posts or in general. These summaries are not necessarily literal summaries of other posts, but more abstract summarization of skills. For example, after an OP was convinced by many debunkers that the evidence of Pizzagate was fake, a debunker summarized how checking the validity of the evidence should be the basic first step to identify suspicious information:

“Whenever seeing something suspicious or conspiracy theories, the first essential step is to confirm and check the basics, that is, is there evidence independent of conspiracy sources that confirms the person even wrote/said/did what is being claimed in the information.”

Another debunker summarized that statements with many statistics involved should always be taken with caution:

“You should always be suspicious with any statistics you see, unless you have a detailed analysis of how those statistics were collected as well as the mathematical toolkit necessary to interpret them. Statistics are ridiculously easy to misrepresent, manipulate, and misinterpret, either intentionally or not, and as such always deserve a skeptical eye.”

4.3.6 Social interactions.

While debunking posts mostly focused on the information itself, users also interacted with each other socially, which was often triggered by the debunking discussions. Social interactions differ from the previous types of community activities among the debunkers mainly because such social interactions are not directly about the debunked information itself but triggered by them. The most common type of social interaction is appreciation and gratitude. Users expressed such emotions when a debunker posted a comprehensive, well-researched, and persuasive debunking post. For instance, an OP appraised a debunker for the convincing and factual debunking explanation:

“I really love seeing this kind of debunking. Takes the claim and focuses on evaluating its own merits.”

Even though Reddit allows users to upvote or downvote to express their emotions toward posts, many users still wrote replies to express their appreciation. Some also complimented the quality of the debunking explanation and said they would share it with their family members and friends. The debunkers that received appreciation and gratitude from the poster and other users may be motivated to keep providing high-quality debunking explanations because of the positive community feedback.

Another common type of social interaction is reminding community norms. Oftentimes debunking involves discussions of political ideologies and sensitive topics and may trigger abusive languages and attacks toward individuals. Not only the community moderators but also other users would remind community norms to maintain a healthy and friendly community environment. For example, in a thread discussing political protests, a debunker attacked the OP, without providing any actual debunking explanations. Another debunker reminded him/her of the community rule: *“Attack the claim, not the person.”* The debunker quickly apologized for the impolite attack toward the original poster: *“Sorry. I forgot where I was.”*

4.4 Original posters (OPs) interactions with debunking responses

Next, we present the themes that emerged from OPs' responses after they received debunks from the community. Overall, OPs did not commonly comment on whether they perceived a debunk to be effective. Only 15% of the coded threads have OPs explicitly expressing their perceived efficacies of the debunking answers.

4.4.1 Prevalence of OPs' interactions.

Among 108 threads that we coded, OP joined the discussion in 48% (52 out of 108) of the coded threads, indicating OPs' involvement in the interactions of debunking and managing the thread. In our study sample, OPs' interactions attracted more reflections and responses from community

members. Threads with OP's involvement received 27.36 replies in average, which was larger than their counterpart without OPs' interactions (number of average replies = 23.44)

4.4.2 *Types of OPs' interactions with debunking responses.*

In addition to the general interaction patterns such as *request*, *summarization* and *social interactions* as shown in Section 4.3, two unique types of OP's interaction emerged that well captured this special role's motivations and characteristics, including *clarification* and *questioning*.

Clarification was commonly adopted by OPs to better contextualize the debunking question and facilitate the debunking process. For example, an OP clarified the motivations to ask for debunking, and supplemented important backgrounds and details that might be useful to generate effective debunking replies:

"I came across this on my YouTube recommendations and it seemed interesting. He uses a map created by Plato and states that "Atlantes" is on the same place on the map as the Eye of the Sahara is in Africa. Now all this seems hard to believe, especially when the map does not show Atlantes in the same place as the Eye of the Sahara. He also cites user comments and seems to bring up the same "evidence" multiple times to hammer it into your brain... I'm interested to see what you come up with!"

Such clarifications were surprisingly common in the replies instead of the original question descriptions and appeared to be interactive which aimed to call for community members' participation in the debunking discussion. Some OPs also clarified their personal judgment with explanations, elucidating specific points and providing initial directions for debunking:

Here's what I find shady in this story: (1) the gang members accepted being pictured and filmed; (2) they traveled in a no-plates car, on main roads, before arriving in a no-cellphone area; (3) guys even said that they're in contact with heavy weaponry dudes; (4) everyone went home without ever being threatened. Some voices said that it might have been a sting op by the Romanian Intelligence. What's your take?

Clarification is not only observed in threads where the original post is brief and later clarification is necessary but also those with detailed and specific debunking requests. OPs actively clarify things that may confuse the debunkers as more users are involved in the thread. For example, when an OP realized most of debunkers focused on the wrong direction for debunking, a clarification was made, *"I can believe that dogs are able to identify trace smells. My question is if this is a legitimate investigation technique. How useful is a dog's sense of smell in an actual non-laboratory situation?"*.

Questioning the approach or argument of debunkers was another typical interaction pattern that OPs used to manage the thread and improve the debunking efficacy. Different from requesting supplementary explanations from debunkers, OPs directly expressed their concerns on whether the debunking was valid. For example, when a debunker tried to debunk the question by revealing the lack of supporting material, the OP showed the skepticism, *"The lack of actual report is certainly a reason to criticize the journalism here but it's not a refutation of the statements in the article."* When another debunker listed the judgment with strong supporting evidence, the debunker was convinced and expressed appreciation, *"Sweet thank you! Now we are getting somewhere"*. This example also indicated OP's preferences for specific debunking methods. When they were skeptical about the effectiveness of debunking, they managed the thread by interacting with debunkers and raising doubts till the debunking became comprehensive and convincing.

4.5 An illustrative case of the community debunking process

In the online community, debunking a post often involved many of the previously presented components and activities. In this section, we show an illustrative example of how these activities and interactions are pieced together among the community members. To protect the privacy of the community members, we did not use a real thread from the community as an example. Instead, we provided a sample case that is typical based on our analysis of many real threads in the community.

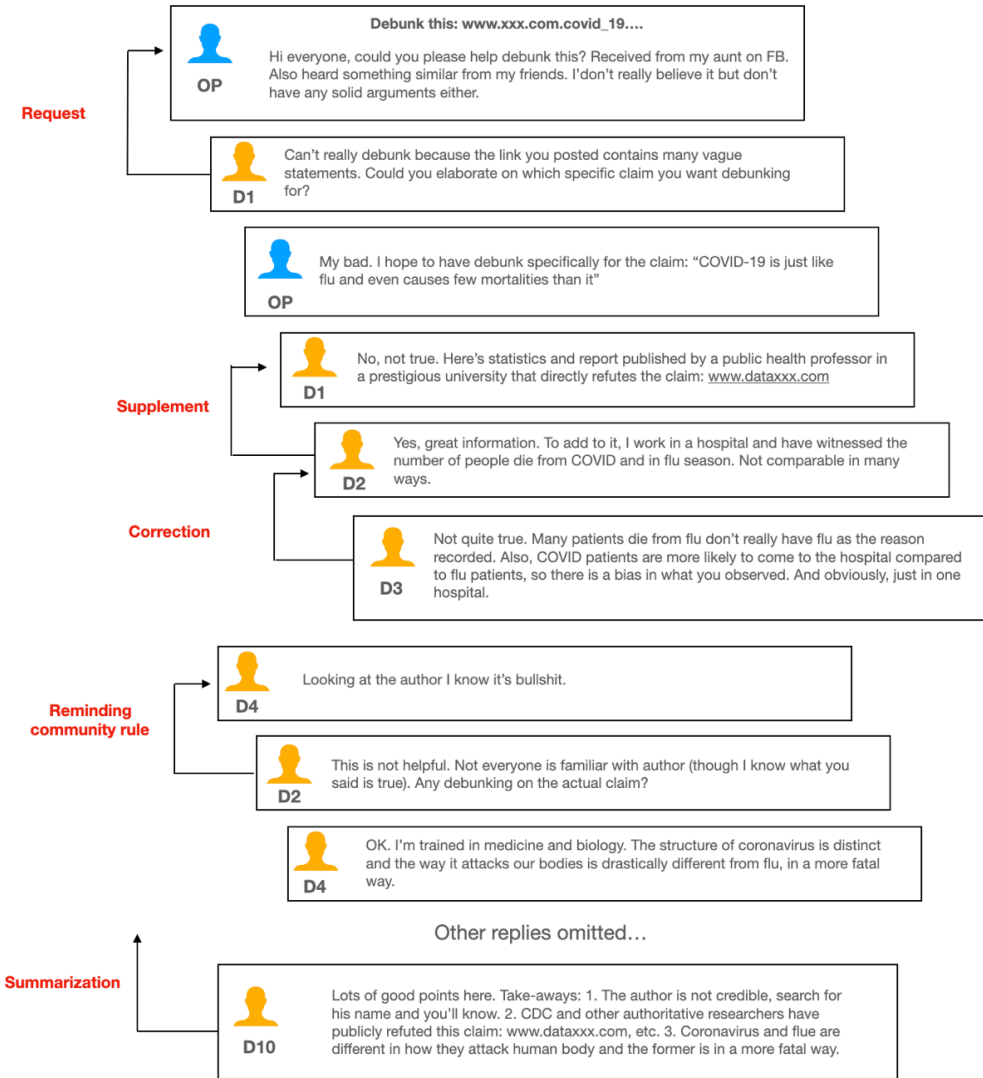


Fig. 2. An illustration of a typical debunking process.

As shown in Figure 2, the OP sought debunking about a piece of information regarding COVID-19 received from family members and friends. D1 offered to help but needed clarification about which specific aspects the OP needed debunking for because the title only included a link to a blog. Otherwise, D1 found the information be not debunkable. The OP responded by clarifying that the

OP needs debunking for the claim that “*COVID-19 is just like flu and caused even less mortalities.*” With the clarification, D1 quickly responded by providing a debunking explanation that the claim is incorrect, based on statistics and interpretations from a credible source who is a university professor. Agreeing with D1 that the claim is incorrect, D2 supplemented D1's explanation by providing personal experience gained when working in a hospital. D3, however, found D2's reply to be incorrect by pointing out potential biases in personal experiences. D4 joined the thread but instead of providing debunking explanations, D4 only pointed out that the author of the article was not credible. D2 reminded D4 of community rules that an actual debunk to the claim should be made.

With many other debunkers providing their debunking explanations, the claim received comprehensive debunking from different aspects. D10 summarized the major take-aways based upon these debunkers' explanations.

This illustrative case shows that, debunking online information, even when the information was relatively easy to debunk and a consensus was reached quickly, users still adopted multiple strategies to improve the explanation and interact with each other (e.g., reminding community rules). Though the debunking goal is to identify whether the information is correct or not, the debunking process enacts more than just labeling it as right or wrong.

5 DISCUSSION

Through a qualitative study of an online community dedicated to information debunking, we revealed that debunking information online in a community setting is complex and often not straightforward. To accomplish the task, individuals and the community collaboratively took on complex work to clarify what to debunk, decide whether the information is debunkable, contextualize the information provided, check information credibility, explain it to other community members, and resolve disagreements. The results complement existing literature on information debunking and community-based debunking systems by providing a nuanced and detailed description of how people individually and collaboratively debunk information in a natural, online setting. We discuss how the study results broaden our understanding of debunking beyond a goal with binary, right-or-wrong outcomes, but instead as a collaborative, social, and learning process that is supported through community-based work. Based on the results, we provide design implications for community and crowd-based debunking systems and suggest that they should support the collaborative, social, and learning aspects of debunking, as well as recognizing the labor of crowd and community debunkers.

5.1 Community-based debunking: beyond right or wrong

Through unpacking the individual and collaborative work taken on by the community members when debunking information online, we broaden the concept of debunking that is not limited to a goal with binary, right-or-wrong outcomes. The result of this study indicates that, in a natural, community-based online environment, debunking is more than just labeling information as credible or not. Instead, debunking initiates collaboration among community members where they actively negotiate with each other, supplement and correct others' explanations, and request more information. In addition, they interact socially in the debunking process, such as expressing gratitude and reminding community rules. Further, some debunkers and even the OPs value the debunking process as a learning and educational opportunity, through which they can summarize

and disseminate high-level debunking skills and tips to the community (Section 4.3.5). In this section, we discuss the three extended dimensions of community-based debunking processes.

5.1.1 Debunking as a negotiated and collaborative effort.

First, we found that instead of a linear and individual process, community-based debunking is often negotiated and collaborative both among the debunkers and between debunkers and the OPs. Before debunkers delved into the debunking process such as searching for external information and leveraging domain knowledge, they decided whether the information presented is debunkable. If not, they negotiated with the OP to modify and clarify in the post to make it debunkable. After debunkers provided their debunking explanations, the OPs may also negotiate with the debunkers to ask for further clarifications. While debunkers had their own opinions and strategies to debunk information, it is common that a post was debunked through collaborative efforts, instead of disjoint, individual efforts. For instance, debunkers supplemented others' explanations using additional external sources, even when previous explanations have been accepted. By supplementing other debunkers' explanations either on different aspects of the information or using different debunking strategies, a post received extensive and comprehensive debunking explanations that are more than enough to decide its credibility. Debunkers also pointed out mistakes in others' explanations and correct them. In addition, some OPs provide contexts of the information such as where it was circulated by whom, which helped other debunkers better understand the information and may potentially improve the quality of debunking. For example, some OPs clarified that the information was circulated in their family's Facebook group and the members are mostly older adults. The debunkers then provided less statistical debunking answers that are tailored for the OP to share with the family group. Therefore, the final debunking explanations were produced by iterative negotiation and collaboration among the community members, instead of individual, disjoint efforts.

5.1.2 Debunking as a social and community-building process.

Second, community-based debunking involves not only deciding if a post is credible or not. Debunking also serves as an opportunity for community members to interact socially and form and reinforce community rules and values. For example, we observed in our study sample that not only community moderators, but also ordinary users would engage in reminding others of community rules. When debunkers mocked the OPs, other users would oppose the debunkers and requested them to focus on debunking and not attacking others. When debunkers simply identified a piece of information as false but provided no explanations, other debunkers requested further explanations. Forming and reinforcing community rules are significant in the debunking process, as the rules are constantly brought up by community members to remind others and ensure that the debunking explanations are of good quality. In addition, the appreciation and gratitude from other users are likely to serve as positive social feedback that may encourage and promote more high-quality debunking from the community.

Similarly, what we have observed as a social and community-building process confirms findings from previous literature in information crowdsourcing tasks such as Wikipedia. In collective information evaluation, for example. *r/ChangeMyView*, most studies still focused on individual users' responses and effectiveness to persuade others. The debunking literature also rarely views the debunkers as a community and leaves out how the debunkers form, apply, and reinforce the community rules and social norms they developed [6,12]. Understanding the social and community-building process can provide insights into designing to support the debunkers as a community.

5.1.3 *Debunking as a learning and educational process.*

Third, our analysis shows that community members often viewed and valued debunking not only as a goal to identify incorrect information and claims, but also as a process through which they can learn about and teach others how to debunk information in general. This is manifested in cases when some posts were debunked easily and a consensus was reached quickly, debunkers still posted detailed, in-depth, and comprehensive explanations that aimed not only to debunk the specific posts, but also summarize high-level debunking skills. In addition, when there are debunkers who simply posted a short reply that even though identified the post as incorrect but provided no reasoning, other debunkers often suggested them to provide detailed explanations because they believed that the platform is not only for identifying what is wrong but also provides an opportunity for those to learn about how to assess information credibility in general. The potential of community-based debunking for peer-based learning and education resonates with recent literature. For example, educational news literacy training to help laypeople identify misinformation achieved no effect [2]. Instead, it was found that even simple corrections of misinformation from peers can improve people's ability to identify misinformation [3]. Zeng et al. also observed that debunking posts by individuals received fewer criticisms than those by police official accounts [62]. On the contrary, debunking by official accounts received more doubts about its validity and requests for more information. It is possible that debunking provided by peers and other community members often has the chance to be interactive, and they can have most or even all their doubts and questions resolved, thus increases their trust toward the debunking explanation.

While we recognize that the collaborative, social, and learning processes we observed in this study have been reported and discussed in many other studies of online communities and social media on different topics [18,38,54], we note that they are less explored and acknowledged in crowd and community debunking literature, which often regards online debunking as an individual and objective endeavor, rather than a negotiated, collaborative, social, and learning process. We believe extending our understanding of debunking beyond the objective, linear, and individual dimension sheds light on better designing community and crowd-based debunking systems that recognize and support the collaboration and interactions which naturally arise among community debunkers in real-life, natural environments, as we describe below.

5.2 **Design implications for community-based debunking systems**

Based on our results, we provide design implications for designing community and crowd-based debunking systems that received increasing attention and efforts from social media platforms.

5.2.1 *Designing for supporting negotiation and collaboration in debunking.*

As our study results demonstrate, in real-life, natural online environments, information debunking is rarely straightforward and individual. Instead, debunkers often negotiated with each other and with users who believed the information is true to decide what specific aspects to debunk, how to best debunk it, and how to improve existing debunking explanations. Therefore, online debunking systems should support not only individual debunking functions such as labeling information as incorrect or submitting notes, but also collaborative debunking functions, such as supplementation, correction, and request. These collaboration functions, as demonstrated in our study results, can enhance the quality of debunking explanations by incorporating input from a wider range of community members and allow them to actively build upon, extend, and correct others' debunking

posts. These observations may inform community-driven debunking systems such as Twitter's Birdwatch [66]. It was found that only less than half of the debunkers submitted notes to explain their reasoning of information debunking, which raises concerns around the quality of community-based debunking [64]. Even though Birdwatch implements functions that allow users to report others' debunking as "low quality" (e.g., lacking evidence, biased), such functions are still very basic and do not allow richer interactions such as supplementation, correction, and request, which are common in community debunking as observed in our study. Implementing these functions could potentially enable community members to enhance the quality of others' debunking collaboratively.

5.2.2 Designing for supporting learning and knowledge accumulation as a community.

We found that besides identifying information as false or not, debunkers also value the debunking process as an opportunity to disseminate high-level debunking meta-knowledge and skills. In addition, the summarization of high-level skills and previously posted high-quality debunking explanations are often referenced and mentioned in other posts. These summarization and high-quality posts can be seen as knowledge accumulated collectively as a community. Therefore, we suggest that the system should be designed to support learning in the debunking process by tagging and highlighting summarizations of meta-knowledge and high-level skills from the crowd. In addition, previous high-quality debunking explanations can be automatically recommended to similar posts, as it represents valuable knowledge accumulated through community work. Even if a post is eventually determined to contain misinformation and be taken down, the knowledge accumulated in the process of debunking it should persist and be preserved instead of thrown away. We also argue that this may also apply to other crowdsourced information tasks such as Wikipedia projects and collective sensemaking, where knowledge generated as a community, directly or indirectly related to the goal, should be valued and kept as the products of community work. Possible design considerations include archiving and publicizing high-quality and collectively contributed discussion and summaries for the ease of future references. Lastly, we note that there are other platforms or subreddits that may host similar discussions. It is also our future work to analyze data from these platforms to validate our study findings.

5.2.3 Designing for recognizing the labor of crowd and community debunkers.

Finally, as our study results demonstrate, the work involved in debunking online information is varied and complex. Debunkers often needed to negotiate with each other multiple times to request clarification before they can start debunking. They assessed surface characteristics of information, searched for credible external information sources, and leveraged personal knowledge. They also enhanced others' debunking or improved their own based on community members' requests and suggestions. Unpacking the varied and complex work taken on by community debunkers points to the labor that is often neglected in previous literature, which primarily focuses on the outcome of debunking. Therefore, we suggest that the system should recognize and give credits to crowd and community debunkers, who often voluntarily engage in debunking work that is complex, tedious, and time-consuming. We warn of the risk of simply treating community debunkers' voluntary work as free labor. Community-based debunking systems can recognize debunkers' work by means such as awarding badges to debunkers and listing debunkers' usernames when posts they debunked were tagged as misinformation and taken down.

5.3 Limitations and future work

Our study has several limitations. First, we studied only one online community whose members may not represent the whole spectrum of online information debunkers. The subreddit has established rules and norms, which may affect the interactions among its community members and are distinct from other online communities and social media platforms. Second, our study only analyzed online discussions, without probing into actual users' experiences and concerns. In future work, we plan to conduct interviews with users who have helped debunk online information to understand the actual experiences, and investigate how they take measures, such as sharing the key take-aways outside the community, to raise the impact of debunking. Third, our study did not quantitatively examine the prevalence of different individual and collaborative debunking strategies and which ones were most effective in persuading others. This is primarily due to the difficulty in manually annotating a larger sample of the data and comprehensively coding the debunking strategies, as well as reliably inferring the effectiveness of the debunking answers perceived by the users. It is our future work to quantitatively analyze how different debunking strategies and interactions among community members contributed to debunking qualities and experiences perceived by other users. We will use the themes qualitatively derived from this preliminary study to guide future computational studies and scale up the analysis. In addition, several findings from this study are based on qualitative observations and may not be generalizable. It is our future work to quantitatively assess the validity and generalizability of these findings.

6 CONCLUSION

We conducted an exploratory, qualitative analysis of an online community dedicated to information debunking. We reported the individual and collaborative work involved in information debunking. We found that community-based information debunking is rarely linear and individual but often negotiated, collaborative, and social, where users actively supplemented, corrected, requested, summarized, and interacted socially around others' debunking explanations. Users also valued debunking as a learning and educational opportunity to distill and disseminate high-level debunking skills. We discuss how our study extends our understanding of debunking and design implications to design community and crowd-based debunking systems that support collaborative, social, and learning processes and recognizes the labor of crowd and community debunkers.

REFERENCES

- [1] Jairo L. Alves, Leila Weitzel, Paulo Quaresma, Carlos E. Cardoso, and Luan Cunha. 2019. Brazilian Presidential Elections in the Era of Misinformation: A Machine Learning Approach to Analyse Fake News. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (Lecture Notes in Computer Science)*, Springer International Publishing, Cham, 72–84. DOI:https://doi.org/10.1007/978-3-030-33904-3_7
- [2] Sumitra Badrinathan. 2020. *Educative Interventions to Combat Misinformation: Evidence From a Field Experiment in India*. (August 2020). DOI:<https://doi.org/10.33774/apsa-2020-1nbcl>
- [3] Sumitra Badrinathan, Simon Chauchard, and D J Flynn. An Experiment on Fact-checking Misinformation in India. 44.
- [4] Michael Barthel, Amy Mitchell, and Jesse Holcomb. 2016. Many Americans Believe Fake News Is Sowing Confusion. Pew Research Center's Journalism Project. Retrieved July 11, 2021 from <https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>
- [5] Md Momen Bhuiyan, Amy X. Zhang, Connie Moon Sehat, and Tanushree Mitra. 2020. Investigating Differences in Crowdsourced News Credibility Assessment: Raters, Tasks, and Expert Criteria. arXiv:2008.09533 [cs] (August 2020). DOI:<https://doi.org/10.1145/3415164>
- [6] Jiang Bian, Yunpeng Zhao, Ramzi G Salloum, Yi Guo, Mo Wang, Mattia Prosperi, Hansi Zhang, Xinsong Du, Laura J Ramirez-Diaz, Zhe He, and Yuan Sun. 2017. Using Social Media Data to Understand the Impact of Promotional

- Information on Laypeople's Discussions: A Case Study of Lynch Syndrome. *J Med Internet Res* 19, 12 (December 2017). DOI:<https://doi.org/10.2196/jmir.9266>
- [7] Petter Bae Brandtzaeg and Asbjørn Følstad. 2017. Trust and distrust in online fact-checking services. *Commun. ACM* 60, 9 (August 2017), 65–71. DOI:<https://doi.org/10.1145/3122803>
- [8] Petter Bae Brandtzaeg, Asbjørn Følstad, and María Ángeles Chaparro Domínguez. 2018. How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. *Journalism Practice* 12, 9 (October 2018), 1109–1129. DOI:<https://doi.org/10.1080/17512786.2017.1363657>
- [9] Juan Carlos Alvarez de la Vega, Marta E. Cecchinato, and John Rooksby. 2021. Why lose control? A Study of Freelancers' Experiences with Gig Economy Platforms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (May 2021), 1–14.
- [10] Alton Y.K. Chua, Sin-Mei Cheah, Dion H. Goh, and Ee-Peng Lim. 2016. COLLECTIVE RUMOR CORRECTION ON THE DEATH HOAX. *PACIS 2016 Proceedings* (June 2016). Retrieved from <https://aisel.laisnet.org/pacis2016/178>
- [11] Juliet Corbin and Anselm Strauss. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 3rd ed. Sage Publications, Inc, Thousand Oaks, CA, US. DOI:<https://doi.org/10.4135/9781452230153>
- [12] Ziv Epstein, Gordon Pennycook, and David Rand. 2020. Will the Crowd Game the Algorithm? Using Laysperson Judgments to Combat Misinformation on Social Media by Downranking Distrusted Sources. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, Association for Computing Machinery, New York, NY, USA, 1–11. DOI:<https://doi.org/10.1145/3313831.3376232>
- [13] Kiran Garimella and Dean Eckles. 2020. Images and Misinformation in Political Groups: Evidence from WhatsApp in India. *arXiv:2005.09784 [cs]* (May 2020). Retrieved July 7, 2021 from <http://arxiv.org/abs/2005.09784>
- [14] Sarah A. Gilbert. 2020. "I run the world's largest historical outreach project and it's on a cesspool of a website." Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 019:1-019:27. DOI:<https://doi.org/10.1145/3392822>
- [15] L. Graves and F. Cherubini. 2016. The Rise of Fact-Checking Sites in Europe. *Reuters Institute for the Study of Journalism*. Retrieved July 10, 2021 from <https://ora.ox.ac.uk/objects/uuid:d55ef650-e351-4526-b942-6c9e00129ad7>
- [16] Kathleen Gray and Jacinta Tobin. 2010. Introducing an online community into a clinical education setting: a pilot study of student and staff engagement and outcomes using blended learning. *BMC Medical Education* 10, 1 (January 2010), 6. DOI:<https://doi.org/10.1186/1472-6920-10-6>
- [17] Andrew M. Guess and Benjamin A. Lyons. 2020. Misinformation, Disinformation, and Online Propaganda. In *Social Media and Democracy: The State of the Field, Prospects for Reform*, Joshua A. Tucker and Nathaniel Persily (eds.). Cambridge University Press, Cambridge, 10–33. Retrieved July 2, 2021 from <https://www.cambridge.org/core/books/social-media-and-democracy/misinformation-disinformation-and-online-propaganda/D14406A631AA181839ED896916598500>
- [18] Xinning Gui, Yubo Kou, Kathleen H. Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, Association for Computing Machinery, Denver, Colorado, USA, 4520–4533. DOI:<https://doi.org/10.1145/3025453.3025891>
- [19] Aaron Halfaker, R. Stuart Geiger, and Loren G. Terveen. 2014. Snuggle: designing for efficient socialization and ideological critique. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, Association for Computing Machinery, New York, NY, USA, 311–320. DOI:<https://doi.org/10.1145/2556288.2557313>
- [20] Thomas Heverin and Lisl Zach. 2012. Use of microblogging for collective sense-making during violent crises: A study of three campus shootings. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 34–47. DOI:<https://doi.org/10.1002/asi.21685>
- [21] Kyle Hunt, Bairong Wang, and Jun Zhuang. 2020. Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: a case study on shelters and ID checks. *Nat Hazards* 103, 1 (August 2020), 861–883. DOI:<https://doi.org/10.1007/s11069-020-04016-6>
- [22] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 048:1-048:27. DOI:<https://doi.org/10.1145/3392854>
- [23] Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, S. M. Murshid Hasan, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, Abrar Ahmad Chughtai, and Holly Seale. 2020. COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *Am J Trop Med Hyg* 103, 4 (October 2020), 1621–1629. DOI:<https://doi.org/10.4269/ajtmh.20-0812>
- [24] Shagun Jhaver, Pranil Vora, and Amy Bruckman. 2017. Designing for Civil Conversations: Lessons Learned from ChangeMyView. *Georgia Institute of Technology*. Retrieved November 15, 2021 from <https://smartech.gatech.edu/handle/1853/59080>

- [25] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08), Association for Computing Machinery, New York, NY, USA, 37–46. DOI:<https://doi.org/10.1145/1460563.1460572>
- [26] Yubo Kou and Colin M. Gray. 2017. Supporting Distributed Critique through Interpretation and Sense-Making in an Online Creative Community. Proc. ACM Hum.-Comput. Interact. 1, CSCW (December 2017), 60:1-60:18. DOI:<https://doi.org/10.1145/3134695>
- [27] Yubo Kou and Xinning Gui. 2020. Mediating Community-AI Interaction through Situated Explanation: The Case of AI-Led Moderation. Proc. ACM Hum.-Comput. Interact. 4, CSCW2 (October 2020), 102:1-102:27. DOI:<https://doi.org/10.1145/3415173>
- [28] Yubo Kou and Xinning Gui. 2021. Flag and Flaggability in Automated Moderation: The Case of Reporting Toxic Behavior in an Online Game Community. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–12. DOI:<https://doi.org/10.1145/3411764.3445279>
- [29] Yubo Kou, Xinning Gui, Yunan Chen, and Kathleen Pine. 2017. Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis. Proc. ACM Hum.-Comput. Interact. 1, CSCW (December 2017), 1–21. DOI:<https://doi.org/10.1145/3134696>
- [30] Amanda Lazar, Norman Makoto Su, Jeffrey Bardzell, and Shaowen Bardzell. 2019. Parting the Red Sea: Sociotechnical Systems and Lived Experiences of Menopause. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Association for Computing Machinery, New York, NY, USA, 1–16. DOI:<https://doi.org/10.1145/3290605.3300710>
- [31] Jonathan Lazar, Gabriele Meiselwitz, and Jinjuan Feng. 2007. Understanding web credibility: a synthesis of the research literature. Found. Trends Hum.-Comput. Interact. 1, 2 (January 2007), 139–202. DOI:<https://doi.org/10.1561/1100000007>
- [32] Alex Leavitt and John J. Robinson. 2017. The Role of Information Visibility in Network Gatekeeping: Information Aggregation on Reddit during Crisis Events. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17), Association for Computing Machinery, New York, NY, USA, 1246–1261. DOI:<https://doi.org/10.1145/2998181.2998299>
- [33] Young Ji Lee, Charles Kamen, Liz Margolies, and Ulrike Boehmer. 2019. Online health community experiences of sexual minority women with cancer. J Am Med Inform Assoc 26, 8–9 (July 2019), 759–766. DOI:<https://doi.org/10.1093/jamia/ocz103>
- [34] Jianing Li. 2020. Toward a Research Agenda on Political Misinformation and Corrective Information. Political Communication 37, 1 (January 2020), 125–135. DOI:<https://doi.org/10.1080/10584609.2020.1716499>
- [35] Yuyang Liang. 2017. Knowledge Sharing in Online Discussion Threads: What Predicts the Ratings? In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17), Association for Computing Machinery, New York, NY, USA, 146–154. DOI:<https://doi.org/10.1145/2998181.2998217>
- [36] Tzu Tsun Luk, Shengzhi Zhao, Xue Weng, Janet Yuen-Ha Wong, Yongda Socrates Wu, Sai Yin Ho, Tai Hing Lam, and Man Ping Wang. 2020. Exposure to health misinformation about COVID-19 and increased tobacco and alcohol use: a population-based survey in Hong Kong. Tobacco Control (August 2020). DOI:<https://doi.org/10.1136/tobaccocontrol-2020-055960>
- [37] Tim K. Mackey, Vidya Purushothaman, Michael Haupt, Matthew C. Nali, and Jiawei Li. 2021. Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter. The Lancet Digital Health 3, 2 (February 2021), e72–e75. DOI:[https://doi.org/10.1016/S2589-7500\(20\)30318-6](https://doi.org/10.1016/S2589-7500(20)30318-6)
- [38] Lena Mamykina, Drashko Nakikj, and Noemie Elhadad. 2015. Collective Sensemaking in Online Health Forums. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), Association for Computing Machinery, Seoul, Republic of Korea, 3217–3226. DOI:<https://doi.org/10.1145/2702123.2702566>
- [39] Amanda Menking and Ingrid Erickson. 2015. The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), Association for Computing Machinery, New York, NY, USA, 207–210. DOI:<https://doi.org/10.1145/2702123.2702514>
- [40] Amy Mitchell, Jeffrey Gottfried, Galen Stocking, Mason Walker, and Sophia Fedeli. 2019. Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed. Pew Research Center's Journalism Project. Retrieved July 2, 2021 from <https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>
- [41] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. Tweeting is believing? understanding microblog credibility perceptions. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12), Association for Computing Machinery, New York, NY, USA, 441–450. DOI:<https://doi.org/10.1145/2145204.2145274>

- [42] Mohsen Mosleh, Cameron Martel, Dean Eckles, and David Rand. 2021. Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–13. Retrieved July 12, 2021 from <https://doi.org/10.1145/3411764.3445642>
- [43] Elena Musi, Debanjan Ghosh, and Smaranda Muresan. 2018. ChangeMyView Through Concessions: Do Concessions Increase Persuasion? arXiv:1806.03223 [cs] (June 2018). Retrieved November 26, 2021 from <http://arxiv.org/abs/1806.03223>
- [44] Sabine Niederer and José van Dijck. 2010. Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society* 12, 8 (December 2010), 1368–1387. DOI:<https://doi.org/10.1177/1461444810365297>
- [45] Katherine Ognyanova, David Lazer, and Matthew Baum. THE STATE OF THE NATION: A 50-STATE COVID-19 SURVEY. 22.
- [46] Gordon Pennycook and David G. Rand. Fighting misinformation on social media using crowdsourced judgments of news source quality [Psychological and Cognitive Sciences]. Proceedings of the National Academy of Sciences of the United States of America. DOI:<https://doi.org/10.1073/pnas.1806781116>
- [47] Liza Potts and Angela Harrison. 2013. Interfaces as rhetorical constructions: reddit and 4chan during the boston marathon bombings. In Proceedings of the 31st ACM international conference on Design of communication (SIGDOC '13), Association for Computing Machinery, New York, NY, USA, 143–150. DOI:<https://doi.org/10.1145/2507065.2507079>
- [48] J Hunter Priniski and Zachary Horne. Attitude Change on Reddit's Change My View. 6.
- [49] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society* 7, 2 (April 2021), 20563051211019004. DOI:<https://doi.org/10.1177/20563051211019004>
- [50] Miriam Redi, Besnik Fetahu, Jonathan Morgan, and Dario Taraborelli. 2019. Citation Needed: A Taxonomy and Algorithmic Assessment of Wikipedia's Verifiability. In The World Wide Web Conference on - WWW '19, ACM Press, San Francisco, CA, USA, 1567–1578. DOI:<https://doi.org/10.1145/3308558.3313618>
- [51] Jodi Schneider, Alexandre Passant, and John G. Breslin. 2011. Understanding and improving Wikipedia article discussion spaces. In Proceedings of the 2011 ACM Symposium on Applied Computing (SAC '11), Association for Computing Machinery, New York, NY, USA, 808–813. DOI:<https://doi.org/10.1145/1982185.1982358>
- [52] Haeseung Seo, Aiping Xiong, and Dongwon Lee. 2019. Trust It or Not: Effects of Machine-Learning Warnings in Helping Individuals Mitigate Misinformation. In Proceedings of the 10th ACM Conference on Web Science (WebSci '19), Association for Computing Machinery, New York, NY, USA, 265–274. DOI:<https://doi.org/10.1145/3292522.3326012>
- [53] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A Platform for Tracking Online Misinformation. In Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion), International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 745–750. DOI:<https://doi.org/10.1145/2872518.2890098>
- [54] Nischal Shrestha, Titus Barik, and Chris Parmin. 2021. Remote, but Connected: How #TidyTuesday Provides an Online Community of Practice for Data Scientists. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 52:1-52:31. DOI:<https://doi.org/10.1145/3449126>
- [55] Kumar Bhargav Srinivasan, Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Chenhao Tan. 2019. Content Removal as a Moderation Strategy: Compliance and Other Outcomes in the ChangeMyView Community. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (November 2019), 163:1-163:21. DOI:<https://doi.org/10.1145/3359265>
- [56] Wen-Ying Sylvia Chou and Anna Gaysynsky. 2020. A Prologue to the Special Issue: Health Misinformation on Social Media. *Am J Public Health* 110, S3 (October 2020), S270–S272. DOI:<https://doi.org/10.2105/AJPH.2020.305943>
- [57] Yannis Theocharis, Will Lowe, Jan W. van Deth, and Gema García-Albacete. 2015. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society* 18, 2 (February 2015), 202–220. DOI:<https://doi.org/10.1080/1369118X.2014.948035>
- [58] Zeynep Tufekci. 2017. Twitter and Tear Gas. Yale University Press. Retrieved July 13, 2021 from <https://www.degruyter.com/document/doi/10.12987/9780300228175/html>
- [59] Emily K. Vraga and Leticia Bode. 2017. Using Expert Sources to Correct Health Misinformation in Social Media. *Science Communication* 39, 5 (October 2017), 621–645. DOI:<https://doi.org/10.1177/1075547017731776>
- [60] Dustin Wright and Isabelle Augenstein. 2020. Claim Check-Worthiness Detection as Positive Unlabelled Learning. In Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 476–488. DOI:<https://doi.org/10.18653/v1/2020.findings-emnlp.43>

- [61] Liang Wu, Fred Morstatter, Kathleen M. Carley, and Huan Liu. 2019. Misinformation in Social Media: Definition, Manipulation, and Detection. *SIGKDD Explor. Newsl.* 21, 2 (November 2019), 80–90. DOI:<https://doi.org/10.1145/3373464.3373475>
- [62] Jing Zeng, Jean Burgess, and Axel Bruns. 2019. Is citizen journalism better than professional journalism for fact-checking rumours in China? How Weibo users verified information following the 2015 Tianjin blasts. *Global Media and China* 4, 1 (March 2019), 13–35. DOI:<https://doi.org/10.1177/2059436419834124>
- [63] Yuehua Zhao, Jingwei Da, and Jiaqi Yan. 2021. Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management* 58, 1 (January 2021), 102390. DOI:<https://doi.org/10.1016/j.ipm.2020.102390>
- [64] 2021. Analysis: Twitter's crowdsourced fact-checking experiment reveals problems. Poynter. Retrieved July 7, 2021 from <https://www.poynter.org/fact-checking/2021/analysis-twitters-crowdsourced-fact-checking-experiment-reveals-problems/>
- [65] 2022. Wikipedia:Purpose. Wikipedia. Retrieved April 25, 2022 from <https://en.wikipedia.org/w/index.php?title=Wikipedia:Purpose&oldid=1064118274>
- [66] Introducing Birdwatch, a community-based approach to misinformation. Retrieved June 22, 2021 from https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation
- [67] Twitter's "Birdwatch" Aims to Crowdsourc Fight Against Misinformation. NPR.org. Retrieved June 22, 2021 from <https://www.npr.org/2021/02/10/965839888/twitters-birdwatch-aims-to-crowdsourc-fight-against-misinformation>
- [68] 微博发布 2018 年辟谣报告：这三条谣言被举报最多--社会·法治--人民网. Retrieved July 11, 2021 from <http://society.people.com.cn/n1/2019/0201/c1008-30606166.html>
- [69] PSAW: Python Pushshift.io API Wrapper (for comment/submission search) – PSAW 0.0.12 documentation. Retrieved July 12, 2021 from <https://psaw.readthedocs.io/en/latest/>

Appendix A

Community rules: Posting rules

ALL post titles use the following format: “Debunk This:[main claim here]” Posts Must include (in the body of your post):

- An explanation of what you would like debunked.
- At least one source.
- For video links, time stamps to specific parts that you want debunked.

* Topic should be at least 2 months old. This information will give other users a specific point to look into, and will increase the chances people replying to your post.

Community rules: General rules

All posters and commenters must remain civil, refrain from abusive language and hominem attacks, and must act in good faith; Obvious trolling will be removed and repeat offenders may be banned. No self-promotion or advertising. No NSFW/NSFL images in posts or comments. Descriptions are fine. Mods reserve the right to remove low-effort posts. Each post should be about wanting to debunk something, a story of debunking, or an example of debunking a myth. Meta posts are allowed.

Appendix B

Codebook of individual work of information debunking

Category	Definition	Example
Debunking based on specific characteristics of the information	Debunkers decide the credibility of the information based on specific characteristics of the information itself, such as whether it is accompanied with supporting evidence and whether it has been cherry-picked.	<i>“The video looks like it’s been edited and slowed down. If you look closely, there’s a weird dark mark that doesn’t look natural in an unedited video.”</i>
Debunking based on external information	Debunkers reference and highlight evidence from external information sources such as research papers, mainstream news outlets, and posts from other subreddits that either directly debunk the same information or provide useful information.	<i>“It has already been debunked many times elsewhere. Here’s one: https://www.politifact.com/factchecks/2020/may/06/blog-posting/dont-fall-conspiracy-about-dr-anthony-fauci-hydrox/”</i>
Debunking with personal knowledge	Debunkers decide the credibility of information based on knowledge obtained through personal professional training or experiences that are gained first-hand.	<i>“The numbers are very close. I know this because I work as a nurse in ICU.”</i>
Debunking with multiple strategies	Debunkers utilize more than one debunking strategy to decide the credibility of the information	<i>“First, take a look at this:https://www.businessinsider.com/coronavirus-lab-manmade-myth-debunked-2020-6 The details of the virus’ genome make it extremely unlikely to have a laboratory origin. Then there’s the lack of that person’s credibility, given that she only published scientific papers on a different subject.”</i>

Codebook of community work of information debunking

Category	Definition	Example
Supplementation	Debunkers add details, explanations, and reasoning to others’ debunking answers to improve the rigor and quality of the debunking.	<i>“Also, the claim of curing cancer is a huge red flag. There are many varying types of cancer and thousand different ways of treating them.”</i>
Request	Debunkers and OPs ask for additional information, details, evidence, or explicit reasoning to help them debunk.	<i>“Could you let me know the context? Circumstance of abortion? Trimester?”</i>
Correction	Debunkers and OPs spot mistakes or deficiencies in other debunkers’ answers and provide correction to fix them.	<i>“It’s not an estimation but actually real data out of a study population of 260,000 inmates in the US.”</i>
Summarization	Debunkers and OPs synthesize key take-aways and high-level messages from others’ debunking answers and provide concise summaries for the community.	<i>“Major takeaway points are: 1. The radio wave emitted by 5G is non-ionizing; 2. the only non-ionizing radiation which could be damaging is microwave, but that’s been deliberately altered to do that exact task in controlled devices; 3. It is indeed listed by the WHO as a possible carcinogen, but on the same list as caffeine.”</i>
Debate	Multiple debunkers hold different opinions and in order to resolve, they exchanged personal judgment, provided corresponding evidence, and uncovered the other’s flaws (if any) in the nested replies	Example omitted due to the contextual nature of this category. Please refer to Section 4.3.4 (page 15) for example.
Social interactions	Debunkers express appreciation or gratitude or remind others of community rules, which do not contribute to the debunking directly.	<i>“I really love seeing this kind of debunking. Takes the claim and focuses on evaluating its own merits.”</i>

Received: July 2021, Revised: November 2021, Accepted April 2022.