Folk Models of Misinformation on Social Media

Filipo Sharevski
DePaul University
faharevs@depaul.edu

Amy Devine
DePaul University
adevine@depaul.edu

Emma Pieroni
DePaul University
epieroni@depaul.edu

Peter Jachim
DePaul University
pjachim@depaul.edu

***Abstract*—In this paper we investigate what *folk models of misinformation* exist on social media with a sample of 235 social media users. Work on social media misinformation does not investigate how ordinary users deal with it; rather, the focus is mostly on the anxiety, tensions, or divisions misinformation creates. Studying only the structural aspects also overlooks how misinformation is internalized by users on social media and thus is quick to prescribe "inoculation" strategies for the presumed lack of immunity to misinformation. How users grapple with social media content to develop "natural immunity" as a precursor to misinformation resilience, however, remains an open question. We have identified at least five *folk models* that conceptualize misinformation as either: *political (counter)argumentation*, *out-of-context narratives*, *inherently fallacious information*, *external propaganda*, or simply *entertainment*. We use the rich conceptualizations embodied in these folk models to uncover how social media users minimize adverse reactions to misinformation encounters in their everyday lives.**

I. Introduction

Misinformation refers to false or inaccurate information that is disseminated regardless of intent to mislead [84], [34]. When the dissemination is deliberate so to mislead the information consumers, it is referred to as *disinformation* [88]. Misinformation and disinformation—often materialized as fake news, rumors, conspiracy theories, trolling, hoaxes, etc.—undermine the *integrity* of public information-sharing systems because they introduce alternative narratives that counter the verifiable factual information shared from credible sources [67]. In the past, dissemination and sharing of information was limited to media outlets, press, and publications with editorial control responsible for preserving the factual integrity of information [2], making it hard for misinformation to spread fast and reach large amount of people [62].

The expansion of the public information-sharing systems to encompass social media platforms, however, morphed misinformation into a much larger threat to the integrity of public information. Now people, as information consumers, gained the ability to disseminate, share, as well as produce, information in droves, rendering editorial fact-checks impractical and platform intervention as intrusive to the information sharing in a public discourse form [6]. Alternative misinformation narratives, attached to polarizing topics like elections, public health issues, and foreign affairs [74], quickly flooded social media platforms, exposing information consumers to amplified falsehoods on a daily basis [7]. Individual pieces of misinformation might not have been a problem for information consumers to reject, but repeated exposure created familiarity with the falsehoods that, research suggests [20], [45], is more important for people than the objective truth.

Without rigorous platform oversight on social media and with increasingly selective editorial fact-checks in traditional media [6], [54], this so-called "illusory truth effect" of misinformation turned into a direct threat to the liberal democratic order through elections interference [81], undermined the public health through fear mongering [11], and increased social anxiety through heightened interpersonal polarization [68], [77]. Misinformation, unleashed as such, drew the attention of the scientific community to devise strategies for limiting the "illusory truth effect" on a user-level [65], [35], [23], [5], [43], and work on detection, measurement, and containment of falsehoods on a system-level [69], [29], [55], [58], [50].

On social media, both lines of work concentrate on preserving the high-level integrity of the shared information. One user-level strategy is "prebunking" and includes a forewarning as well as a preemptive refutation of the falsehoods [35], [37]. Another strategy is the use of "accuracy nudges" [47] or simply encouraging people to think about accuracy of questionable information on platforms. If these strategies do not prevent users from believing misinformation, "debunking" is the next step where verifiable corrections of the falsehoods from credible sources are presented in order to break the illusion of truth [52], [16]. When users refuse to heed the "experts'" corrections, they can still counter misinformation by leveraging the "wisdom of the crowds" [3].

On a system-level, several misinformation curbing strategies exist [14]. The first focuses the analysis on the misinformation *content* by leveraging natural language processing techniques or authoritative information sources [42]. Another focuses the analysis on the misinformation *context* instead by exploring the interplay between end users, publishers, and articles [63], [33], [50], [58]. A third one focuses the analysis on the misinformation *creation* by using automation to detect, report, and remove accounts that disseminate misinformation [55], [29]. If the misinformation still cannot be curbed, platforms have the option for algorithmic moderation by either obscuring it with warnings covers or attaching warning labels [18], [61].

While the coordinated scientific response helps curb misinformation [46], all of the strategies work in a *top-down* fashion, i.e. they start with a fixed definition and set of assumptions about misinformation and then work down what *users* or *systems* can do about it. Little attention, however, is devoted to what "definition" or "asumptions" users have while

Network and Distributed System Security (NDSS) Symposium 2023
27 February - 3 March 2023, San Diego, CA, USA
ISBN 1-891562-83-5
https://dx.doi.org/10.14722/ndss.2023.24293
www.ndss-symposium.org

encountering and processing misinformation on social media. In other words, hardly any work so far has studied what mental processing models exist in dealing with misinformation from a folk perspective in the first place, and how well these models serve a user in a *bottom-up* fashion. Identifying these *folk models of misinformation* is an important contribution because they could provide a more nuanced account about the natural human response to misinformation by users as information consumers, and as such, improve the coordinated effort for curbing misinformation.

The paradigm of *misinformation folk models on social media*, thus, is the main focus of our work as an effort to explore how people respond to misinformation through contextualization, interpretations, and various forms of actions. The inductive articulation of the folk models of misinformation also serve a bigger purpose in approaching misinformation as an adverse situation, an information-based threat not just to systems but to people as information consumers. The conscious acknowledgement of adverse situations force people to decide how to response to misinformation, learn from experiences, and develop strategies for dealing with emerging (mis)information. The outcome, as much as it descriptively uncovers the current misinformation folklore, is an essential starting point in learning how people are predisposed to build *resilience to misinformation*. This knowledge, in turn, could benefit the already underway top-down mass "inocculation" strategizing for misinformation [35], [30], [37].

Through a study with 235 social media users in the United States, we found out that misinformation is conceptualized in several distinct ways based on how the information on social media relates to the facts known to the public. Looking top-down, one would expect that most, if not all, of the participants in our study will possess what we named the *inherently fallacious information* folk model – misinformation is any information *unfaithful* to to known facts, regardless of contexts or intentions [62]. However, the most prevalent folk model was the *political counter(argumentation)* – misinformation was conceptualized as any information with faithfulness to *selective* facts relative to political and ideological contexts, created and disseminated with a political agenda-setting or argument-winning intentionality.

The second most prevalent folk model referred to all misinformation as *out-of-context narratives* with a *questionable* faithfulness to known facts due to selection of improbable alternative contexts, created and disseminated with speculative intentions. In the political counter(argumentation) and the out-of-context narratives folk models, facts *do appear* as part of the misinformation, but the appearance is predicated to a certain extent by self-promoting goals of other users on social media, instead of misinformation being only the *disappearance* of facts caused by foreign actors. The foreign actors enter the picture with the *external propaganda* folk model, where misinformation is any information with a *fluctuating* faithfulness to known facts relative to shifting contexts, created and disseminated with a propagandistic intentions. We found that the users also employed an *entertainment* folk model where misinformation is assumed as any information with a *tangential* faithfulness to known facts relative to humorous or sarcastic contexts, usually shared with entertaining intentions.

As to who the "misinformers" are, the participants in our sample did not just point at the foreign actors; they blamed the other "uneducated, ignorant, gullible, bigoted, deluded, shameless, and insistent on being stupid" users for spreading misinformation. The purpose of misinformation was not necessarily to mislead a user on social media, but to serve as political ammunition, stir the pot, increase profit, and entertain. When the participants in our sample were exposed to misinformation on social media, they employed variants of the dual model of information assessment proposed in [39] that incorporates both *analytical* processing (e.g. reference to scientific data, fact checking) and *heuristic* processing (e.g. gut feelings, linguistic formatting inconsistencies).

To report on the findings from our study, we review the top-down approach for curbing misinformation in Section II as synergistic to the bottom-up that is in the main focus of this paper. Section III places the *folk models of misinformation* in a broader context of folk models of security and information consumption on social media. Section IV provides the methodological details of our study and Sections V, VI, VII, VIII, and IX elaborate each of the five folk models in great details with a support of participants' quotations. Section X analyzes the response of misinformation in regards the actions users employ to immunize themselves from misinformation on social media and Section XI traces the evolution of folk models from security and social media participation to misinformation. Section XII provides an additional dimension on how each of the folk models assesses misinformation with the implications of the folk models for future misinformation "inoculation strategies" discussed in Section XIII before we conclude the paper in Section XIV.

II. Responding to Misinformation

A. User Responses to Misinformation

Exploring *why* people fall for fake news in general, Pennycook et al. [49] point to several behavioural factors, namely, lack of careful reasoning and relevant knowledge, use of heuristics such as familiarity, and inattention that associated with internalizing and spreading misinformation. Social media users also just ignore the fake news they come across on Facebook or Twitter according to Tandoc et al. [17], unlikely to reply and attempt to refute a misinformation post [73]. Colliander [13] identifies a "herd immunity" where users were found to be more resilient to misinformation when leveraging crowd refutations to falsehoods on social media.

When debunking was present as an intervention against misinformation, users' response was contextual to the topic and participation structure. Fact-checking, used as a correction to false information on Reddit, was less welcomed on partisan subreddits than on general ones like r/politics [44]. For a period, it was believed that corrections frequently fail to reduce misperceptions among the targeted political ideological groups or that debunking actually creates a "backfire effect" in which corrections actually increase misperceptions about misinformation in general [41]. Despite some indication that this effect is another threat to information integrity and might transfer also on social media [12], [59], Wood and Proter [83], Swire-Thompson et al. [70], and Kirchner et al. [32] found that the "backfire effect" is not as serious as initially thought, with further evidence suggesting that users do heed misinformation corrections and platform warnings [61].

As to *who* are those that are more drawn to misinformation, studies point out that misinformation on social media is entrenched in the partisan agenda and users battle with alternative narratives on almost every issue [82], [48], [51]. It was initially suggested by Bhatt et al. [9] that the proclivity for alternative narratives is most salient among the right-leaning users, but new user evidence from the alt-platform Gettr reveals that misinformation is equally attractive for left-leaning or moderate users because it "keeps them abreast with the latest argumentation of the right online" [60].

B. *System Responses to Misinformation*

While corrections, "accuracy nudges," and fact-checking might work against misinformation, they are slow and difficult to scale to the pace of information sharing on social media platforms. Very often, misinformation claims are promulgated and spread unchecked for periods of time, prompting platforms to turn to a system automation for a rapid response to such *content*. From a natural language processing perspective, such a system first has to determine which claims require verification, then to find sources supporting or refuting the claim, and finally to classify the claim as misinformation or not [25]. In determination, systems employ metadata features such as the number of likes and re-posts [89] or content features (e.g. word length and repeated words) [28] to detect rumors and misinformation candidates. Finding sources for checking the veracity of these candidates depends on manual fact checking input, through automatic means of "stance detection" [56] and "stylometric identification" help in this step, especially with human-generated text [57]. Finally, platforms have the option to apply more granular classifications with several degrees of truthfulness [4].

Anti-misinformation systems also focus on the *context* and infrastructure that supports dissemination of falsehoods on social media. Partisan-biased publications, low-crediblility users, and history of fake news articles are considered in a classification system that gives an early alert of potential fake news within a small time delay of the original news postings [50], [63]. The early warning of misinformation on social media could also be supplemented by accelerated veracity checks that leverage inputs from crowd workers with different levels of expertise [33]. Through causality and structural metrics between the users [58] as well as infrastructural features (e.g. domain registrations, certificates, or web hosting configurations) [26], the contextual detection could also determine the "spreading" accounts and the websites from which these spreaders introduce the misinformation on social media.

A combination of content and context that additionally leverages participatory traits such as commenting, same-title posting, and bursts of account creation is another strategy for automated detection of misinformation behaviour on social media [55]. Hashtags, mentions, likes and engagement elements such as volume and length of posts do also help with detection of misinformation as they reveal patterns of disinformation dissemination from so-called "troll farms" [27], [29], [28]. Cross-platform coordination and targeting of innocuous accounts in polarised discourses is possible in cases where misinformation and falsehoods are aggressively utilized to cause personal harms to individuals [69].

Misinformation on social media does not always come in human generated, text-only form and systems do need to factor multimedia claims or "memes" as well as text/audio/video "deepfakes" [38], [86]. As memetic content spread across multiple platforms, tracing the virality and influence is imperative of automated systems in spotting and flagging falsehoods that embed offensive and hate speech [85]. By supplementing the detection of misinformation memes with facial recognition, character extraction and factoring the meme image template, automated systems could trace the evolution of memes as well as the emergence of new meme genres [8], [71].

Textual "deepfakes" detection systems use probabilistic language modeling to determine if, first, a claim in a social media post is machine generated, and second, the sense of the falsehoods generated [19], [22]. Finally, systems for audiovisual "deepfakes" leverage affective cues such as eye movements and tone of the voice to help social media platforms separate misinformative content from original [40]. It is important to mention that the application of automation alone could lead to non-negligible false positives and false negatives and, at this point, social media platforms do need a human intervention in rendering the final misinformation moderation decision [24].

III. Folk Models of Security and Social Media Participation

Identifying the folk models of misinformation on social media is predicated on contextualization and various forms of actions regarding computer security in general and social media in particular. The low comprehension of security that lead to many early exploits, prompted the scientific community to explore what mental models users employ when they make security-related decisions [78]. Users dealt with security, as suggested by Camp [10], by employing one of the five mental models: *physical*, *criminal*, *medical*, *warfare* and *market* models. Each model contextualizes security either through lenses of damage to computers, a breach in a computer, the possibility of "infectious" weakening of computers, the attack on computers, or failure to protect computers.

When non-expert users take expert security advice, users either did not conceptualize viruses as security exploits, thought of them as a *buggy* software (the *buggy* model), or identified them as written by *mischievous* individuals (the *mischief* model) or *criminals* (*criminal* model) according to Wash [79]. Hackers were seen as either *opportunistic* criminals looking for financial data (the *burglar* model) *high profile targeting* criminals (the *bigfish* model), or young, technically oriented individuals lacking moral restraint (the *vandal* model) that sometimes seek financial reward (the *contractor* model).

Exploring users' concepts of how computer security is subverted and how they are affected, Wash and Rader [80] found that users believe that malicious software *originates* on Internet, *creates* visible problems, and one could reasonably *protect* from malware (by careful surfing or use of protection software). As for who are targets of malware, users believe hackers spare neither *ordinary* nor *high-profile* users. When making security decisions beyond using just protection software, users only declaratively support "strong" security but nonetheless *reuse* passwords, contrary to experts' advice [75].

Exploring the participatory mental models on social media, Johnson et al. [31] found that users, while being able to control their privacy, are nonetheless concerned with inappropriately sharing content with members of the friend network (*insider threat* model). Users on social media too were less concerned about hackers' breaching their passwords and more concerned about who can see and comment on their posts as well as tag their accounts [76]. The proclivity towards a private circle participation is not surprising given that users often regret their posts (for reasons such as: revealing too much, direct criticism, direct attack, or blunder) and try to repair their stance [64].

IV. Folk Models of Misinformation on Social Media

A. Research Questions

Developing an understanding of what mental models people actually possess is not to generalize a population, but rather to explore a phenomenon in depth, as pointed out by Wash in the seminal work on folk models of home computer security [79]. To this goal, we focused on five questions respective to how users deal with misinformation on social media:

1) *Folk Models*: How do social media users model misinformation (i.e. what is misinformation according to them)?
2) *Origins*: Who do social media users believe creates and benefits from disseminating misinformation?
3) *Purpose*: How do social media users interpret the purpose of misinformation?
4) *Response*: What actions social media users take in responding to misinformation?
5) *Assessment*: How do social media users decide whether given content is or is not misinformation?

B. Sample

Our study was approved by the Institutional Review Board (IRB) of our institution before we fielded a survey protocol with semi-structured, open-ended questions, listed in the Appendix. After the survey segment, participants were given the option for a follow-up contact using voluntarily provided e-mails where they could expand, if they wished, more on the questions of the survey. We sampled a population who were 18 years or older, from the United States, regularly use social media, and have encountered misinformation on social media platforms. We used Amazon Mechanical Turk to recruit our participants as it provides the means to pre-screen participants for social media on the mainstream platforms (e.g. Twitter, Facebook, Reddit, etc.) and create custom screenings for other alternative platforms (e.g. Gab, Gettr, and Parler).

We performed a preliminary pilot test with 30 participants in which the responses started to show the contours of at least three distinct impressions of misinformation. We therefore decided to recruit at least an order of magnitude bigger sample size in order to a) get richer description of each folk model, and b) ensure we do not miss any possible mental models that could exist outside of the small pilot sample. From data we collected in the main study, we identified and removed low quality responses where participants used their answers to provide dishonest commentary unrelated to the questions of the study (e.g. persistently use insults and derogatory terms against the research team). We ended with a sample of total of 235 participants. The participation in the study was anonymous and allowed users to skip any question they were uncomfortable answering, took around 40 minutes to complete it, and participants were offered a compensation rate of $6.20 each. The demographics are given in Table I. A total of 64 participants opted for a follow-up after we completed the survey data collection, and their answers were incorporated into the overall data analysis.

TABLE I: Sample Demographic Distribution

Gender				
Female	**Male**	**Prefer not to say**		
102 (43.4%)	117 (49.78%)	16 (6.82%)		
Age				
[18-30]	**[31-40]**	**[41-50]**	**[51-60]**	**[61+]**
32 (12.76%)	100 (42.55%)	60 (25.53%)	28 (11.91%)	15 (6.38%)
Political leanings				
Left	**Moderate**	**Right**	**Apolitical**	
115 (48.93%)	61 (25.95%)	49 (20.85%)	10 (4.25%)	
Misinformation encountered on:				
Facebook	**Twitter**	**Reddit**	**4chan**	**Other Plaforms**
156	131	54	17	6

C. Method and Analysis

In the pilot test we initially asked the participants to provide their take on misinformation, but the answers were drifting towards editorial policies (e.g. Fox/NYT) and curatorial decisions (e.g. users' feeds) instead of the misinformation's nature itself. To ensure validity to the task of identifying the misinformation mental models in depth, we decided to introduce the participants in the main study to the generalized definition of misinformation on social media proposed by Wu et al. [84] as indicated in the interview protocol in the Appendix. Participants were then asked to characterize misinformation in their own terms and where on social media they have encountered claims that they considered misinformation. Next, participants were asked for their opinions on where misinformation comes from, what is its purpose on social media, and who creates it and benefits from it. Lastly, participants elaborated how do they suspect a certain social media post is misinformation, and what tactics they employ when dealing with misinformation.

The qualitative responses were coded, using the codebook given in the Appendix, and categorized in respect: a) salient features that generally conceptualize misinformation; b) origins of misinformation; c) misinformation purpose; and d) assessment of misinformation claims. Two independent researchers analyzed the raw responses received, achieving a strong level of inter-coder agreement (Cohen's $\kappa = .86$), averaged over the entire set of study questions. For the response to misinformation, we used the social media engagement affordances available (e.g. comment, mute, block, report, etc.) as well as the ability to fact-check certain claims [72], [4].

We utilized a thematic analysis methodology to identify the themes and sub-themes most saliently emerging from the responses in our sample. The themes were summarized to describe individual folk models, and for each folk model, we created sub-themes that describe the origins of misinformation, the purpose of it, and how individuals respond to misinformation. In reporting the results, we utilized as much as possible verbatim quotation of participants' answers, emphasized in "*italics*" and with a reference to the participant as either **PXYZ#** or [**PXYZ#**], where **P** denotes **participant**, **X** denotes the **number** of the participant in the sample (ordered by the time of participation), **Y** denotes their **gender** identity (**F** - female, **M** - male, **N** - preferred not to say), **Z** denotes their **political** identity (**L** - left-leaning, **M** - moderate, **R** - right-leaning; **A** - apolitical), and **#** denotes the upper bound of their **age bracket**. For example, **P136FL60** refers to **participant 136**, **female**, **left-leaning**, **age bracket [51-60]**.

V. Political (Counter)Argumentation

A. Folk Model of Misinformation

The predominant conceptualization of misinformation within our sample refers to any information that has faithfulness to *selective* facts relative to political and ideological contexts, created and disseminated with agenda-setting or argument-winning intentionality. This folk model moves the social media users from passive consumers to active partakers that shape misinformation with competing interpretations based on the rhetoric of political personas, commentators, think-tanks, and vocal supporters. In the effort to "*avoid admitting political defeat at any cost*" as **P64MM40** put it, the intentionality of the political argumentation shifts from massive public opinion manipulation to "*hammering faulty logic into selection of facts as long as that discredits the 'other' side and ultimately wins an argument*" [**P14FR40**].

Participants' concept of misinformation as "*political statements that exaggerate the party agenda*" [**P169FL50**] is driven towards misleading suggestions about a preferred political stance, rather than towards promulgation of absolute falsehoods and inaccuracies [84]. Participants acknowledge that "*falsehoods, inaccuracies, and unsourced claims could appear*" [**P32ML50**] in the the political (counter) argumentation, but they are neither exclusive nor always "*outlandish*" [**P115FR40**], as was the case in past misinformation campaigns targeting political issues [87]. Users do not necessarily think of *fatefulness* to known facts [62], but instead favoring "*selective facts and incomplete evidence*" [**P86FL50**] when other users "*piece together their opinion and put it on social media*" [**P30FL30**]. The responses, in general, uncover an entitlement to one's own opinion *and* one's own facts concept of misinformation on social media [36], fostered by the need for "*reinforcement of one's political beliefs*" [**P58FL50**].

B. Origins of Misinformation

The political counter(argumentation) as misinformation posits a deliberate *selection* of facts that fit into one's political inclinations. The "*people with different political perspectives wanting their beliefs validated*" [**P105FL40**], in this folk model, are considered to be the originators of misinformation on social media. These people "*select facts in accordance with their selfish agendas*" [**P116ML60**] and constantly "*try to convince other people there point of view is correct*" [**P121MM50**]. The reason for bringing misinformation to social media is that "*these people are made in the image of the political leaders they flock to*" [**P13FR50**].

The participants in our sample described the users from the "other side" in the political counter(argumentation) as: "*uneducated, bigoted, prevaricating, and shameless hypocrites*" [**P65FL60**], "*truly deluded, insistent on being stupid*" [**P69FL61+**], "*people who refuse to accept reality*" [**P79MM61+**], and "*impulsive, uneducated people*" [**P96FR60**]. It could be very well that these descriptions refer to the vocal social media users that amplify and alternative narratives "*cooked in political echo chambers*" [**P102FL40**] instead of referring the users themselves as misinformation selectors. Either way, the "other side" was culpable of "*willing to overlook truth*" [**P116ML60**] just to come "*off as important and knowledgeable*" [**P70MM50**].

The "*bitterness*" [**P75FR60**], in the mind of our participants "*comes from people that 'know' the information is wrong, but they still like to create drama, so will post it anyway just to see the kind of arguments that will come from it*" [**P94FL50**]. The accusations, thus, do not cast the "other side" as ignorant [32] or lazy [48], but on the contrary as agitated and determined to win an argument. In the words of **P106ML40**, "*these are people wanting their beliefs validated so they seek out facts that confirm them, then latch onto ones that appear to do so, and sometimes articulate and spread them in a political context that turns them into misinformation; I imagine that's more common that outright intentionally telling lies*."

C. Misinformation Purpose

Most of our participants thought that misinformation "*serves political purposes as it incites people to hate political opponents for bogus reasons*" [**P21FL60**]. Misinformation, exploiting people's "*confirmation bias, makes people more politically close minded*," **P212MR40** pointed out. Misinformation, **P44ML30** added, "*creates a 'hive mind' on social media from a group of followers that becomes loyal to a political cause and amplifies the sounds bites*." Our participants also ushered direct accusation against particular political points of contention. One group believed that misinformation's purpose is "*to protect democrat politicians*" [**P95FR40**] "*as repetition of lies and misinformation is a very old liberal tactic of brainwashing the populace*" [**P156MR61+**]. Another believed that misinformation's purpose is to "*harass liberals*" [**P102FL40**] and "*convince people to vote for Donald Trump*" [**P29ML50**]. Presently, some of the participants believed misinformation "*keeps the conservatives occupied, encourages discourse*" [**P229FL40**], but in equal measure "*creates the illusion that left-wing positions are popular*" [**P68MR40**].

D. Response to Misinformation

Seeing political arguments as misinformation, participants opted to *ignore* the "other side" when sensing that either "*mocking and taunting is about to replace a civilized conversation*" [**P220MM40**] or "*the platform will impose it's own abridged definition of 'free speech'*" [**P45MR60**]. However, participants with this folk model switched tactics to *reporting* if the political arguments was seen as "*harmful*" **[P39ML50]** or "*if there's a chance of actual harm resulting from people believing the misinformation*" **[P15ML50]**. The *fact-checking* tactics were also employed by several of the participants, but mostly driven towards resolving any possible doubts. As **P109FL61+** pointed out, they respond "*depending on the source and type of claim(s)—either investigate the truth of the claim or just dismiss it as the creation of a sick mind.*"

Blocking/muting and asking others to do so was a tactic justified by self-preservation means of "*avoiding misinformation trash to not take up space on my timeline*" [**P14FR40**], "*not wasting my time reading it*" [**P129MM40**], or "*not amplifying the other agenda by engaging with it*" [**P181FL40**]. *Reporting*, in the sense of bottom-up resilience, was seen by the participants as "*helping the administrators learn the latest lies on the political spectrum*" [**P32ML50**], "*prevent calls for violence, defamation, and hate to materialize*" [**P220MM40**], and "*stop amplifying foreign bots*" [**P96FR60**].

P164FM60 underlined that "*it is important to be aware of the false information that is spreading to be informed of what some others may be thinking, as being ignorant could lead to another January 6th.*" Many participants found it critically important to talk to their in-laws or senior family members "*point them to the facts and tell them they should stop following/listening to those people on social media*" [**P152MA40**]. Naivety was the vulnerability identified for those social media users lacking extensive heuristics for misinformation assessment or lack of motivation for "*proactive analysis of source credibility*" [**P63MM40**].

VI. Out-of-context Narratives

A. Folk Model of Misinformation

In this folk model, misinformation is any information that has *questionable* faithfulness to known facts due to selection of improbable alternative contexts, created and disseminated with speculative intentionality. "Alternative contexts" refer to narratives around an event or issue at stake that runs counter to the mainstream context and attempts to displace any factual reporting or development around the said event or issue [66], [1]. Participants here conceptualized misinformation as narratives where "*facts with missing, incomplete, or used in a made-up context*" [**P200FM30**] or "*cherry picking events presented out of context in order to support a biased argument*" [**P28ML40**].

Scientific facts or results were also conceptualized as "alternative narratives" as misinformation was also thought of as "*bad research*" [**P190MM30**] or "*misconstrued study results interpreted in limited context*" [**P188MR40**]. The out-of-context narratives could involve political events such as elections, but we classified this line of reasoning as a separate folk model because our participants mostly pointed to misinformation centered around topics such as health (e.g. vaccines, medicinal side-effects), foreign conflicts (e.g. the war in Ukraine) or other issues (e.g. Black Lives Matter, abortion rights, cryptocurrencies, celebrities, influencers).

B. Origins of Misinformation

In the out-of-context folk model, the interpretation of an issue in an alternative context is the catalyst that turns a speculation into misinformation. Tracing the genesis back to its source(s), **P188MR40** explained: "*Misinformation comes, I think often, through selective reading. People want to confirm their narrative, and so they take things out of context, or in limited context. In reality, things are usually more complex. But rather than deal with complexity, simplistic takes that confirm pre-existing narrative biases get read (and shared) more on social media.*" Several others joined to characterize these people as "*well-meaning dummies*" [**P213FA40**]. Misinformation often originated from "*twisting what's actually a personal opinion into one's subjective idea of a fact*" [**P200FM30**] or "*expressing one's emotions as facts (speaking a personal 'truth')*" [**P189ML40**]. This alone might not be sufficient for misinformation to float on social media, but the affordances of the platforms encourage "*innocent garbling, like the 'telephone' game as kids*" [**P171FA50**] to mutate into a speculative narrative.

Many participants accused the "*die hard speculators*" [**P84FM50**] as "*mentally unbalanced*" [**P130ML50**] and "*having poor critical thinking skills*" [**P177ML50**]. Some of the speculators were simply "*craving attention*" [**P33FL40**], some "*just want hear themselves talk*" [**P178FM60**], and some are "*suspicious of everything due to fear, distrust in science, or living in an 'echo chamber'*" [**P188MR40**]. The presumption of innocence is maintained by our participants who think of these social media users as "*normal people who sincerely believe everything they read, not some big bot, troll campaign, or the work of a foreign actor*" [**P159FM40**].

C. Misinformation Purpose

A prevalent conceptualization of misinformation as alternative narratives was to "*stir the pot*" [**P61MM50**], "*muddy the waters*" [**P4FL40**], and "*keep people up in arms*" [**P13FR50**]. Fear was pointed out as one of the ingredients for "*creating extremists and form violent groups of idiots*" [**P8MA40**] and "*riling up people to fight shadows*" [**P54FL40**]. Confusion and doubt we naturally included too, as misinformation was used to "*showing off 'proof' that people are being lied to or misled by the mainstream media.*" [**P12ML40**]. The goal of misinformation, thus, was driven to "*create less trust in institutions*" [**P166NL60**], "*distract from real issues*" [**P33FL40**], and "*directing the consensus to a topic based on hidden agendas*" [**P22ML60**].

Several participants took misinformation to be the main tool for "*indoctrination*" [**P36MM40**] that is ultimately "*designed to cause people to make bad choices*" [**P88ML40**]. Monetary profit was also a purpose that the out-of-context group identified behind misinformation. As **P176MM50** described it: "*Misinformation drives page views; Page views generate money; Follow the money.*" Platforms were accused of "*promoting misinformation posts to boost engagement and their ad revenue*" [**P35FL50**], make people "*spend money unnecessarily*" [**P148FL60**], and as well as to "*pump the price of certain cryptocurrencies*" [**P28ML40**].

D. Response to Misinformation

The participants operating with this folk model were equally keen on fact checking and engaging with the "misinformers" on social media as their "*pledge to refute false assertions*" [**P148FL60**]. Bringing facts in a widely accepted context was more important than simply blocking "misinformers" because, as **P173FA50** reasoned, "*engaging with misinformation posts helps understand where the alternative context comes from and who is behind it*." **P178FM60** strengthened this position asserting that "*it is important to help educate those that listen to such posts; Turning away at least one individual is worth the cost in confronting any deluded 'misinformers'*." Out-of-context narratives, if not directly confronted, "*turn social media into a breeding ground for hatred, racism, misogyny and greed*" asserted **P127FL60**.

Participants with this folk model also possess a considerable degree of self-preservation of their well-being. For example, **P36MM40** explained: *"I have blocked more people since COVID-19 started than the 10 years I have been using Facebook because of speculating nonsense to a point of harassment"*. **P8MA40** added: "*block the people who post misinformation move on - this is the only way to combat it and remain somewhat sane*." Moving on unbothered was premised on the resilience idea that "*the 'misinformers' come to social media to be acknowledged as such*" [**P10FA40**]; most of the misinformers welcome push back not just because they are ready to converse, but because that "*confirms their 'misinformer' identity*" [**P24ML60**].

The out-of-context narrative folk model, unlike the political (counter)argumentation, set some of the participants to reject misinformation in their real life as it was deemed a "*mood killer*" [**P8MA40**] or an "*unnecessary cause to our already high stress level*" [**P178FM60**]. Another reason was to avoid "*making enemies among friends and acquaintances based on difference in interpretation of hot topics*" [**P189FL40**]. The urge for education, especially of senior family members was also present here, especially because "*early in the COVID-19 pandemic the elderly were ignoring the mandates and potentially spreading or dying from the virus*" [**P148FL60**].

VII. Inherently Fallacious Information

A. Folk Model of Misinformation

The inherent fallaciousness for this folk model conceptualizes misinformation as any information unfaithful to known facts, regardless of contexts or intentionality [62]. The two previous folk models acknowledge that misinformation does in many respects involve inaccuracies and falsehoods, but they are driven towards "*wild ideas*" [**P184MM40**] rather than fabrication for fabrication sake. Many participants in this folk model identified misinformation as "*hoaxes that circulate on social media*" [**P91ML60**] that include "*fear mongering information*" [**P169FL50**]. Misinformation as "*lies*" [**P20FL61+**] or "*blatant falsehoods*" [**P157Fl50**] was also mentioned by our participants in reference to "*factually or scientifically incorrect*" information [**P19FM40**].

B. Origins of Misinformation

"*Individual users or bogus news sources*" were mostly blamed for the spread of misinformation on social media [**P23MA40**]. The spread of falsehoods "*comes from ignorance, hate, and anger*" [**P41FL40**] in the view of this group of participants. Participants acknowledged that misinformation could be weaponized for both political and financial gains. As **P99ML61+** pointed out, "*some of [the misinformation is] put out to gain political favor with more extreme voters; Some of it is put out just to get views via controversy*." The spreaders, in the view of **P100NA40** are "*fringe political accounts on Twitter; Mostly extremists like the far-left, or far-right*." Participants characterized the originators and spreaders of misinformation as "*idiots with subpar IQ*" [**P184MM40**], "*ignorant people*" [**P134FL61+**], "*disingenuous people*" [**P57FL50**], or "*opportunists*" [**P38ML60**].

C. Misinformation Purpose

The misinformation purpose, in the view of this group, was "*usually to stir up controversy or sow dissension amongst the masses*" [**P90FL60**]. It exists "*to persuade others into a viewpoint*" [**P23MA40**] and ultimately "*to bring some sort of anarchy or civil disobedience about so as to hurt people*" [**P149MR40**]. Misinformation "*scares and confuses*" [**P57FL50**] by "*manipulating the public opinion*" [**P173ML40**] and ultimately results in "*killing people*" [**P207NL40**]. Here too, participants held the impression that misinformation's purpose is "*to try to get people to click on links*" [**P141FM50**], "*roil them up, and increasing engagement with the platform*" [**P140FL60**]. **P70MM50**, for example, directly pointed out that "*Facebook in particular has been shown to promote misinformation posts to boost engagement on their platform*."

D. Response to Misinformation

The participants operating with the inherently fallacious information model look at misinformation as something that can be thwarted by direct action on social media platforms. *Reporting* and *blocking* therefore were utilized to flag users and misinformation claims to the attention of the platform administrators. As **P23MA40** put it: "*Yes I will report when necessary if the offense is egregious enough*." The best tactics, as explained by **P57FL50**, "*is to always report posts that contain false information, spread hatred or disenchantment; I always block the user as well; I don't want to see that in my feed or give any thought to the poster*." In the view of **P134FL61**, one should "*engage [with the misinformer] if they think there is any chance they can either change their mind, or think posting some counter evidence might get through to others*." While some of the participants noted that there are risks of direct engagement with the misinformation spreaders because "*debating only adds credence to something that deserves no attention*" [**P99ML61+**], the sense of hopelessness and discouragement to directly engage with misinformation was less prevalent than among the other folk models.

VIII. External Propaganda

A. Folk Model of Misinformation

Unlike the previous folk model, the conceptualization of misinformation as "*external propaganda*" [**P8FL40**] refers to information that *fluctuates* its faithfulness to known facts relative to shifting contexts or perceived division-creating intentions. It could be said that the same conceptualization applies to the political (counter)argumentation or the out-of-context narratives, but we consider this one as a separate folk model that has a distinct "*propaganda*" [**P201FL61+**] flavor to the information operations on social media. In the mind of our participants, the external propaganda or *disinformation* was associated "*nation-states*" [**P20FL61+**], but it finds its place among Americans because "*people and organizations that control the media and the government allow to float on social media*" [**P3FR60**].

A distinguishing feature for this folk model is that the decision of whether information is or is it not misinformation is not explicitly in relationships to verifiable facts, but decided based on the actor(s) that produce and disseminate this information. Within this folk model, a piece of information *entirely based on factual evidence* might be considered outside interference and automatically rejected if it appears to be from "*organized propaganda campaigns, meant to undermine the United States*" [**P128ML50**]. Participants in our sample assigned these campaigns both to the "*right-wing media*" [**P20FL61+**] and a "*large bot factory to promote left-wing articles*" [**P60MR40**].

B. Origins of Misinformation

Expectedly, the "'usual suspects" of external propaganda – "*nation states hostile to the United States and her interests*" [**P60MR40**] – were identified as misinformation originators by our participants. **P20FL61+** pulled no punches, pointing to "*Putin, as the Russians have been waging psychological warfare against the west for years, and it's as if we all just pretend it isn't happening.*" **P26FL60**, extended the list of spreaders to include: "*right-wing think tanks, bourgeoisie controlled liberal media, oligarchs, churches, and outside interests trying to destabilize the region.*" The "*bots, sycophants, and complete morons*" were also named as misinformation spreaders [**P120MM40**]. An interesting aspect of this folk model is that it outlines a *chain of misinformation* where: a) "*the fake accounts controlled by the state actors put initial rumors and fabricated facts*" on social media **[P97FR60]**; b) these are picked up and amplified by "*demagogue figureheads that glom onto misinformation that suits their needs*" [**P60MR40**]; and c) kept alive by "*ignorant individuals freely sharing it*" [**P128ML50**] that appropriated the misinformation as the "*preferred truth*" [**P105ML40**].

C. Misinformation Purpose

As a propaganda tool, misinformation was seen as "*serving the function of persuading the more naïve into being fear-mongered into extreme beliefs.*" [**P80FL40**]. In this context, the misinformation was used "*to suppress class consciousness, drive ad revenue, and destabilize the nation*" [**P26FL50**]. A conspirative nature was assigned to misinformation too, as the purpose was seen to "*destroy liberal democracy, rule of law, and replace it with a kleptocratic form of government where the financial elite no longer have pesky regulations or taxes, minorities have no rights, and the enforcement of law becomes arbitrary*" [**P60MR40**]. Misinformation was used by "*bad actors with evil in their hearts*" [**P179MM50**] to exploit "*impulsive people*" [**P96FR60**] and "*idiots*" [**P112FL40**] to "*advance their agenda, for example, erode the trust in institutions*" [**P183ML40**]. The eroded trust together with the increased polarization, caused by propagandist information operations, was seen as a tactic for "*weakening the US economy and military power without fighting a real battle*" [**P55ML40**].

D. Response to Misinformation

Seeing misinformation as an external interference, or simply propaganda, set a subgroup of participants to action as means to counter persuasion. As propaganda messages could fluctuate in regards the faithfulness to known facts [6], these participants prioritized proactive fact checking to "*overshadow it with information with maximum faithfulness to known facts*" [**P225FL30**]. Part of this tactic was demonstrating resistance to "misinformers" and part was educating people in the inner circles, as **P80FL40** put it: "*If I have someone around me who is believing the misinformation about specific hot topics at the moment, I will give them tons of references in hopes I can sway them away from someones' hidden agenda.*"

The participants using this folk model heavily relied on heuristic processing as their "*guts*" **[P189ML40]** were telling them "*not to take emotion-provoking posts at the face value*" [**P80FL40**]. The rejections of emotion-provoking content as a resilience strategy worked for these participants are they were confident in debunking Russian propaganda in particular, "*from Pizzagate to denazification of Ukraine*" [**P96FR60**]. Some of the participants realized that misinformation, in the context of persuasion, could also hinge on religious and political beliefs and felt that experience-based argumentation was needed to counter "*domestic propaganda*" [**P152MM40**].

IX. Entertainment

A. Folk Model of Misinformation

A small, but distinct segment of our sample conceptualized misinformation as any information with a *tangential* faithfulness to known facts relative to humorous or sarcastic contexts, usually created and disseminated with intention to "*entertain*" [**P58FL50**] the social media users. Memes form the largest part of the entertaining misinformation, usually containing "*erroneous statements*" [**P193M30**] and in some cases "*making fun of the misinformation itself*" [**P233FM40**]. The concept of misinformation as entertainment is not to (counter)argue with posts perceived as polarizing, but to "*mock off-the-wall posts*" [**P123FR50**] themselves. Falsehoods contribute to entertainment, if not mocking, for example, in case of "*celebrity rumors*" or "*tabloid content*" [**P133FM61+**] causing a "*laughing reaction*" [**P209FM40**].

B. Origins of Misinformation

Towards the conceptualization of misinformation-as-entertainment, our participants characterized the spreaders as "*people making jokes and other people believing them as real*" [**P233FM40**]. While the notion of "*pranksters making speculation for fun*" [**P18MA61+**] was the main image of the misinformers, it was not lost on this group that memes are used to "*spread a negative view towards a figure, issue, or movement*" [**P58FL50**]. Social media platforms were accused of enabling memes to "*spread like wildfire*" [**P2FL40**], causing misinformation to come to attention to "*people who are naturally drawn to posts with wild and far-fetched ideas*" [**P43MM30**].

C. Misinformation Purpose

Some participants here believed that misinformation "*doesn't have a true function besides satire and entertainment*"

[**P78FL40**]. In the words of **P130ML50** misinformation is a source of entertainment because "*it is fun to watch people describe daily all sorts of impossible stuff that is bothering them.*" The attention seeking element was also mentioned by our participants, suspecting that misinformation is also entertaining to "*users on social media targeting other gullible people that share the misinformation*" [**P59MM40**]. Misinformation-as-entertainment, predominately, showed up in private offline interactions where participants "*held their comments until amongst the friends to laugh together*" [**P130ML50**].

D. Response to Misinformation

While the laughing at misinformation came from a place of self-preservation ("*it's not worth arguing with them, only laugh at them*" [**P216MR50**]), these participants also appear to have an emotional threshold that when crossed, causes them to actively resist by blocking, muting, reporting, or replying. As participant **P111FL50** pointed out, "*these actions are warranted when memes stop being taken humorously and become seeds for more misinformation.*" These participants were inclined to fact check the accuracy behind misinformation but for their own purposes and shared back to the community (both online and offline) when that threshold is crossed. These appeared to be emotionally driven responses such as "*I will sometimes mock how stupid the misinformation is and laugh about it openly*" [**P125ML40**].

X. Immunity to Misinformation

A. (In)attention, fact-checking, and heuristics

Prior work on responding to misinformation identifies several key factors behind the lack of immunity to misinformation: (i) inattention, (ii) lack of careful reasoning, (iii) lack of knowledge, and (iv) use of heuristics such as familiarity for misinformation assessment [49]. Some of our participants justified the inattention as a conscious decision, because by "*not give misinformation the attention that it wants, one lessen the engagement with it and it won't spread as much*" [**P33FL40**]. Participants like **P49FL60** went further in attempting to limit the attention to misinformation by "*downvoting misinformation posts so they don't get more attention*" by other users.

While we found evidence of using familiarity for heuristic misinformation assessment, the participants in our sample were actually careful with possible misinformation on social media. **P8MA40** offered simple, yet careful rule-of-thumb reasoning: "*If it seems insane that is the first clue. Beyond that, I look at the site or the source being used. If there is no source? 90% chance it's a lie. If there is a source/site listed it doesn't take much effort to glance at it and know if it's misinformation or an extremist site*". **P39ML40** also provided a careful reasoning approach: "*If it's a share or post of an article, I look at the source. Is it a news organization, and if so, which one? or is it a blog?. Is the article using 'trigger words,' hostile language, or blanket statements? If yes, most probably is misinformation.*"

While participants in our sample were divisive on the question of journalistic integrity (or lack of thereof) when news reporting was used for obtaining knowledge on a polarized topic, they do resorted to using "*fact checking sites like PolitiFact.org and Snops [sic]*" [**P76ML40**]. Fact checking a post was a tactic that most participants employed when they "*actually cared about the issue*" [**P63MM40**]. In this case, several participants also went "*'googling' the issue*" [**P134FL61+**] or turned to "*reputable sources or official reports*" [**P182FL40**].

Several participants turned to "*scientific evidence*" [**P47ML40**] beyond the fact-checking websites as the felt "*confident to find the truth after years of fine-tuned bullshit detection*" on social media [**P135ML40**]. A small subset opted for a balanced fact checking, reading "*both left- and right-leaning sources*" [**P191MM50**] in addition, and looking for the "*truth somewhere in the middle as journalists are inherently biased*" [**P213FR40**]. Few participants even critiqued the reliance only on secondary fact-checking, as "*people count on mainstream media to tell them what is true or false, and that is wrong*" [**P104MM40**]. Everyone on social media, in the view of **P214FR40** "*should take a piece of information and do background on it themselves to find out what is true and what is fake.*"

Participants in our sample reported accounts in addition to fact checking when they were "*convinced that the account(s) are hostile actors from other countries, bots, or intentionally grifting*" [**P55ML40**]. A group of participants engage in a public refute after they fact-check posts and use "*facts, numbers, charts, quotes—anything that will refute the misinformation items stated*" [**P201FL61+**]. Some of these participants took an interesting approach of overshadowing misinformation with real information as they resorted to "*flood the poster with re-posts, tags, and and replies*" including accurate sources [**P225FL30**]. In their view, showing a direct resistance to the misinformation spreaders by "*publicly calling them out*" [**P136FL60**] is equally important and necessary as it is the "*misinformation cleanup*" [**P27ML40**] the social media platforms should do a better job of.

Prior evidence points that fact-checking, when used as a correction to false information was less welcomed in partisan discussion [44]. We found evidence that the participants employing the political counter(argumentation) folk model do employ fact-checking, and "*immediately try to use the most neutral Google search for the information, to see if it's true or not*" [**P124ML50**]. **P132FM50** clarified that they "*don't even share any possibly polarizing information without checking the facts first.*" As noted previously, here too, the fact-checking is predicated on participants involvement with it. As **P167FR61+** described "*If I really care, I will check out the piece of misinformation through a fact checker, such as FactCheck.org; Or, I will try and verify the information through a different source; If the same information can be found on a reliable source, then there is a good chance it is correct information.*"

B. Ignoring, Replying, Refuting

Social media users, prior evidence suggests [17], ignore the fake news they come across on social media, unlikely to reply and attempt to refute a misinformation post [73]. Our results confirm that "*just ignore misinformation*" is a viable tactic for self-inoculation [**P17FR60**]. One reason, as **P188MR40** indicated, is because "*social media became largely an ineffective place for political discourse, so there is no point to engage with misinformation.*" Another reason is that participants were fearing retribution either from "*the platform as they could* *ban me for refuting it*" [**P3FR60**] or "*being cyber-stalked*" [**P18ML61+**]. Sometimes participants ignored misinformation but only to "*see how much others will refute the post and that itself is a sufficient enough factor to differentiate between misinformation and the truth.*" [**P216MR50**]. Some of our participants just "*laughed and move on*" [**P156MR61+**].

Counter to the evidence of disengagement with misinformation [73], participants in our sample to a large degree were open to engage with the spreaders on social media. Encountering "*inflammatory rhetoric*" [**P140FL60**], participants both replied to refute it and then reported it, "*leaving a comment explaining why this rhetoric is misinformation*" [**P200FM30**]. Many of them refuted misinformation by "*advising the person who posted it that the information is blatantly incorrect and what the correct information is with supporting documentation*" [**P202FL40**]. **P182ML60** reasoned that misinformation spreaders "*'deserve' ad hominem counter-argumentation fusillade dripping with vitriol*" in case "*disproving the the misinformation with facts is futile.*"

Several participants called on the "misinformer" to "*remove the post*" [**P142ML50**] afraid of its inciting a massive rift. Directly challenging the account spreading misinformation by "*engaging in a reasoned debate*" [**P46ML40**] was also a tactic where participants judged that a constructive discourse is possible. Usually, the reply includes "*truthful statements that challenge their misinformation*" [**P22ML60**]. Sometimes, the counter-argumentation was framed in a "shoe on the other foot" metaphor i.e. "*reply with something to make the misinformer think about what they're doing (e.g. what if this is your mom) and the repercussions they might bring to someone*" [**P90FM60**].

The "misinformation spreaders" in participants' circles on social media enabled participants to engage with them to educate others about the perils of misinformation.. **P29ML40** indicated he "*will reply to the person, not trying to convince them of course but rather to help educate others who might be reading the misinformation*" as to "*neutralize the spread of misinformation*" [**P178FM60**]. In **P32ML50**'s view, this tactic "*counters misinformation with as much sourced, unbiased information as possible to the benefit so others could see how easy is to disprove misinformation.*" Our participants are aware that arguing with the "misinformer" leads to "*frustration*" [**P145ML50**] but is worthwhile because it helps participants themselves "*to talk to real people about how to counter it*" [**P107MR40**].

C. Blocking, Muting, Reporting

Blocking, muting, and reporting without fact checking, but instead based on subjective convictions or heuristic assessment as indicated in [49], was also a regular tactic employed in our sample. Some of the participants were quite lenient, giving a "three-strikes-you-are-out" chance for someone on social media to post what they perceived as misinformation before they proceed to "*block, mute, or un-follow and encourage others to do so*" [**P70MM50**]. Block, mute, and then report was a step further taken by participants in our sample, in hope "*the account and the post are taken down*" [**P134FL61+**]. These participants avoided commenting "*lest I accidentally help amplify the post*" [**P20FL61+**].

Some of the participants took an approach where misinformation was only blocked, or reported when deemed "dangerous." **158FL50** indicated that "*if the post itself and the assenters' remarks are like falling into a black hole of insane, then I'm just going to block those people and the poster.*" The reporting was directly to social media admins using the platform affordances. Under "dangerous" our participants considered any misinformation that is "*threat to the public health*" [**P163ML40**] "*civil unrest*" [**P119Fl40**], or *defamatory in nature*" [**P136FL60**]. Some of the participants noted that they, "*after a repeated exposure*" [**P181FL40**] have decided to disengage with a social media platform after giving its admins full hands of work by "*reporting a list of misinformation spreaders*" [**P55ML40**]. Blocking, muting, and un-following was so prevalent in our sample, there were even instances where participants applied to people in their own closest circle: "*Yes, I blocked my own husband on Facebook because he was spreading misinformation regarding the 2020 election; I will block anyone who does this*" [**P53FL40**].

D. Platform Interventions

Respective to platforms intervention from systematic handling of misinformation [50], [63], [25], [89], our participants reckon that "*social media companies should do a better job of removing misinformation, and after repeated violations, banning the offending account*" [**P47ML40**]. Though previous work recommends crowd-sourced interventions against misinformation [33], some of our participants are wary of doing so, as **P3FR60** says: "*I read them, and move on; I can't do anything to stop them; If I were to put an effort into that I could get banned.*" Aware that automated anti-misinformation systems utilize traits such as commenting, same-title posting, and bursts of account creation [27], participants don't want to get involved because they fear their accounts might be banned for "*liking/retweeting the wrong thing*" [**P67MR50**], "*replied with correct information*" [**P993MM40**], or "*claimed otherwise*" [**P220MM40**] than the narratives in polarized posts.

Many participants do offer an approach for platform intervention similar to the ones proposed in the literature for contextual and participatory detection of misinformation [26], [55], [29], [28] as well as cross-platform coordination [69]. A misinformation spreader could be noticed, as recommended by **P100NA40**, "*If the account looks like a bot; For example, it has a bunch of numbers in the handle; The account will also be retweeting a bunch of Q-anon hashtags; I will also notice that the person who posted the tweet contributes to certain fringe websites; They will usually promote their website or substack when spreading misinformation.*" As to what the platform intervention should be, **P91ML60** suggests that "*all social media should have fact checking and anyone posting false information should be banned for life*" and **P144FL50** wishes "*people got a 30 day ban if they share fake news.*"

Helping with the detection and tracing the evolution of memes [85], [8], [71], the systems with human intervention could leverage patterns of participation where the response to platforms' "*censoring arguments, one have to counter in other ways (memes, etc)*" [**P107MR40**]. As the meme spreaders, **P112ML40** suggests that "*enemy nation states would be interested in spreading destabilizing memes as a way to to create rifts, hate, and sow discord in their enemy's backyard.*"

It is worth mentioning that participants in our sample also saw memes as a response to falsehoods and many of them either "*created a funny meme for a post is so crazy that I have to laugh*" [**P216MR50**] or "*posted memes making fun of the misinformation*" [**P173ML40**].

E. Misinformation Moderation

Several of our participants have confirmed they heed misinformation corrections and platform warnings, as previous evidence suggests [83], [32], [61]. **P221MM50**, for example, was "*glad that the fake news that I saw on Twitter was flagged*" and **P55FL40** liked that on Twitter "*sometimes there will be a warning right on it that says it could potentially be misinformation*." Albeit anecdotal, some participants hinted that misinformation corrections failed to reduce misperceptions among the targeted political ideological groups [41], [70]. **P67MR50** pointed out that "*sometimes misinformation is labeled as such; That is not always accurate though*." The "backfire effect" was referred to by **P77MM50** saying that he "*sees misinformation daily from his ultra right-wing friends and even when it gets fact checked by Facebook they still think it is true*." **P96FR40** stated that she "*knows people are warned certain stories that damage democrats are labeled untrue when they were absolutely true*."

XI. Evolution of Security and Social Media Participation Mental Models

Considering the five security mental models proposed in [10], one could trace a cross-pollination, if not explicit evolution, in the misinformation conceptualizations identified in our study. Seen through the political (counter)argumentation lenses, the *physical* mental model is related both to physical damage as "*people have died because of the politicization of a pandemic with misinformation*" [**P65FL60**] as well as "*political reputations damage*" [**P115FR40**] and "*defamation*" [**P220MM40**]. Political counter(argumentation) as misinformation also takes on the *criminal* mental model when pointing to misinformation as being responsible for "*inciting unrest and insurrections in the country*" [**P20FM30**].

The *medical* mental model as an "infectious" weakening the information-sharing systems was hinted in the external propaganda folk model, seeing misinformation as "*weakening the US economy and military power without fighting a real battle*" [**P55ML40**]. The *warfare* mental model too, was contextualized by in regards external propaganda, with the blame placed on the "*the Russians for waging psychological warfare against the west for years*" [**P20FL61+**]. The *market* models as in failure to protect the integrity of information on social media, was referenced by both the out-of-context narratives folk model ("*people not fact checking before sharing*" [**P54FL40**]) and by the inherently fallacious information folk model ("*people who want to believe anything other than what the news media and government provides*" [**P70MM50**]).

Misinformation as a "*buggy* model of information" [79] was seen in each of the folk models as the "bugs" ranged from "*blatant lies*" [**P71ML60**], "*speculations*" [**P123FR50**], "*fabrications or distortions*" [**P63MM40**], "*flawed opinions or logic*" [**P64MM40**], "*garbage*" [**P91ML60**], and "*hate speech*" [**P140FL60**]. Vis-à-vis the spreaders, the financial element was present in all folk models in reference to the *mischief* model ("*people trying to stir up trouble, spread their lies, looking for attention, or looking to get 'likes'*" [**P19FM40**]) and the *burglar* model ("*entities interested in chaos and money*" [**P163ML40**]). The *contractor* model was referenced in the inherently fallacious information folk model as misinformation was seen as steaming from "*disinformation firms*" [**P207NA40**], as well as in the political model where the *"troll farms and Russian bots*" [**P32ML50**] were blamed for the spread of misinformation, as well as in external propaganda where the "*trolls/bots were paid to post it*" [**P96FR60**].

While each of the misinformation folk models have a divergent take on the origins and problems of misinformation [80], [75], they converge on the profile of misinformation targets as social media users held in a high disesteem: "*uneducated, bigoted, prevaricating, and shameless hypocrites*" [**P65FL60**], "*truly deluded, insistent on being stupid*" [**P69FL61+**], "*people who refuse to accept reality*" [**P79MM61+**], "*impulsive, uneducated people*" [**P96FR60**], "*well-meaning dummies*" [**P213FA40**], "*idiots with subpar IQ*" [**P184MM40**], "*ignorant people*" [**P134FL61+**], "*disingenuous people*" [**P57FL50**], and "*gullible people*" [**P59MM40**].

Regarding participatory mental models on social media, the *insider threat* model [31] was mostly driven towards disengagement and suggestive recommendations to family or close friends. **P201FL61+** said that she has "*a few family members I've had to un-follow because they post such ridiculous things*," **P53FL40** said she "*blocked her own husband on Facebook because he was spreading misinformation regarding the 2020 election*," and **P70MM50** gave it "*few chances before un-follow people in his feed*." Suggesting of potential misinformation perils, **P142FM50** shared misinformation posts with her friends and family to "*tell them how silly it is and to be careful*." and **P9ML40** to his parents to "*try to educate them on the correct information*." Also, some participants felt that friends had unsubstantiated claims that "*they are the ones who's got it wrong and are the victim of misinformation*" [**P104MM40**].

The users in our sample, contrary to the evidence in [76], were not concerned about who can see comment on their posts as well as tag their accounts and were quite open take these very same actions to "*counter misinformation with as much sourced, unbiased information as possible*" [**P32ML50**]. We found anecdotal evidence that social media users, when dealing with misinformation, often regret their posts [64]. For example, **P109MR40** revealed that "*I lose control of myself or I'm in a bad mood and I feel the need to make derogatory comments; This is always regrettable*," and **P212MR40** suggested that "*you lose someone [sic] well-intentioned misinformation spreader if you mock or taunt with personal attacks*."

XII. Misinformation Assessment

A. Analytical Assessment

Online information's credibility, according to Metzger [39], is assessed by employing both analytical skills and heuristics in determining whether a claim is a misinformation on social media. Relying on the political counter(argumentation) folk model, **P15FL50** reasoned that "*in a political discussion, especially an argumentative one involving more than one* *perspective, it's likely that at least some misinformation is being spread, and this is often self-evident when one side makes one claim and another makes the opposite, as both claim and counter claim can't be true at the same time.*"

When it came to the out-of-context narrative folk model, participants pointed out that it is "*easy to look for the main context from other reputable sources*" [**P42ML30**] and reject such posts as misinformation. Looking at "*external boogeyman accounts of the Russians containing spelling mistakes, bad grammar, and weird patriotic sounding name*" [**P230MR50**] participants employing the external propaganda folk model conducted in the assessment of misinformation. Participants employing the inherently fallacious information folk model resorted to using "*fact checking sites like PolitiFact.org and Snops [sic]*" [**P167FR61+**], and participants employing the entertainment folk model looked at how "*absurd and stupid*" [**P125ML40**] the misinformation is to the point of laughing.

B. Heuristic Assessment

Participants in our sample also used their "*gut feelings*" [**P114FL50**] to assess misinformation. **P212MR40** pointed that he is wary of misleading titles because "*statistics don't lie but liars use statistics*" in a political (counter)argumentation. **P85FL40** scrutinized "*out of context statements and meme-type images saying something controversial*" for uncovering out-of-context narratives. The participants in the inherently fallacious folk model see a misinformation red flag when "*the claims are usually ridiculous with no supporting facts and posted by untrustworthy sources*" [**P57FL50**]. Participants in the external propaganda folk model used the presence of a clear "*fear-mongering tone*" [**P80FL40**] in claims to determine if they come from bot or troll accounts. In the entertainment folk model, participants followed the approach of "*if it looks too good to be true, it probably is*" [**P90FM60**].

Participants also used "sourcing" cues to figure out who is the "misinformer" on social media. For example, **P50ML40** (political counter(argumentation)), **P100NA40** (inherently fallacious), and **P162ML50** (entertainment) cued misinformation spreaders when there are "*bunch of numbers, generic names, and fake images in the profile, something like Tom87654.*" Participants also used "language" cues as a telltale sign of misinformation. **P69FL61+** (political counter(argumentation)) said she "*probably subconsciously ignore posts with misspellings,*" **P123FR50** (entertainment) avoids posts in which the "*grammar is atrocious,*" and **P171ML60** (out-of-context narratives) recoils when he sees "*bad grammar.*"

The "emotion-check" cue was also employed to discriminate between misinformation and content faithful to known facts. As **P86ML50** (political (counter)argumentation) puts it, "*if something gets an emotional reaction out of you, it is time to question the veracity.*" Emotion-provoking misinformation, emerges on social media when posts "*speak in absolutes and employ either absurd or illogical claims*" [**P4FL40**] (out-of-context narratives). According to **P29ML50** (political (counter)argumentation), if "*the source alone is not a sufficient cue, than, misinformation is easy to pick out because it uses anger, fear, or malice to get the point across.*" Usually, the emotion-provoking misinformation "*contains extreme language that lacks nuance and claims have very strong opinionated comments beneath*" [**P131FL40**].

XIII. Discussion

We introduced the *folk models of misinformation on social media* as an effort to provide additional context to researchers studying misinformation rather than denigrating people who interact with it. There is not a "correct" model of misinformation that could serve as a comparison in the first place, therefore, we avoid giving preference to one model over another. The important aspect of our work is not how *accurate* the model is, but how *well it serves* the needs of a social media user in dealing with misinformation [15]. We presented the folk models as separate conceptualizations but it is important to note that people might have one dominant model they employ aligned with what they felt is the most common type of misinformation, which by no means is exclusive. For example, the out-of-context narratives could be employed in an "*agenda such as politics*" [**P137MM40**] or the external propaganda about "*a vaccine that could have saved thousands of lies has been made to believe it's something evil from the other political party*" [**P201FL61+**] as part of the political (counter)argumentation.

A. Implications

Except the inherently fallacious model, all the remaining ones might appear antithetical to how misinformation is defined among expert communities [84], [67], [85] as misinformation is not "entirely" comprised of falsehoods. This obviously has implications both on the prebunking and debunking efforts. The "accuracy nudges" [47] and the preemptive refutation [35], [37] perhaps have to employ degrees of truthfulness [4] and include more detailed context [61]. Same goes for the debunking with verifiable corrections from credible sources [52], [16] as users on social media are aware that hardly any piece of fake news is entirely false, and hardly any piece of real news is flawless[51].

As intent is a salient element in determining what claims are misinformation on social media, it is important to note that many participants pointed to "bots" as *automated* originators of misinformation. Evidence points to disinformation actors *manually* controlling a large number of accounts to appear legitimate [21], [86], and this misconception should be considered in anti-misinformation efforts. For example, Swire-Thompson's et al. [70] recommendation for designing corrective elements should perhaps also include clarifications of the misinformation origins. It should also be clarified that the state-sponsored disinformation actors often just take controversial matters (e.g. Black Lives Matter, abortion, the women's march) with the goal of luring real users into the discussion and polarizing online opinion [55]. This is important as our analysis suggests that the "other users" on social media are held in high disesteem, which is precisely what state-sponsored disinformation operations are aiming to achieve [87], [55].

B. Ethical Considerations

The purpose of our project was not to generalize to a population; rather, to explore the phenomenon of personal dealing with misinformation in depth. To avoid misleading readers, we did not report definitive numbers of how many users possessed each folk model, nor how the folk models and the accompanying conceptualizations fair with participants' demographics. Instead, we describe the full range of folk models we observed, in a hope that the results can help to elevate the study of misinformation as a whole. One could say that there is a risk of oversimplification where the initial set of folk models of misinformation, expressed on the participants behalf, might not represent the entirety of folk models used to deal with misinformation. We, of course, acknowledge that there are certainly other ways and means that users employ and we welcome every work that brings them to the fore.

C. Limitations

Our research was limited was limited in its scope to U.S. social media users. While Redmiles et al. [53] suggest that Amazon Mechanical Turk responses regarding security and privacy experiences are more representative of the U.S. population, we exercise caution to the generalization of the results as there is little insight into general sampling and sample-related differences when users are broadly queried about misinformation. By asking users directly about how they interact with misinformation, we got a wide variety of insights from a broad range of perspectives. We did not measure the efficacy of these folk models, or the variation between the results of different users' applications of folk models in a myriad of social media settings (users do have a preferred platform, but many of use several platforms interchangeably; also new social media platforms are regularly introduced). We are aware that these folk models represent the contextualization and learned behavior informed by all forms of misinformation that currently exist on social media. Therefore, we are careful to avoid any predictive use of the folk models.

XIV. Conclusion

What misinformation is, or represents, undoubtedly is an evolving concept both in context of security and social media participation. A testimony to this evolution are the five folk models of misinformation identified in this study, which, we underline, are here to help with creating better inoculation strategies to ensure the integrity of the content in information-sharing systems and social media. As such, we hope that we bring an actionable starting point in building bottom-up resilience to falsehoods spread online.

References

[1] J. Albright, "Welcome to the era of fake news," *Media and Communication*, vol. 5, no. 2, pp. 87–89, 2017.

[2] J. Allen, B. Howland, M. Mobius, D. Rothschild, and D. J. Watts, "Evaluating the fake news problem at the scale of the information ecosystem," *Science Advances*, vol. 6, no. 14, p. eaay3539, 2020.

[3] A. A. Arechar, J. N. L. Allen, R. Cole, Z. Epstein, K. Garimella, A. Gully, J. G. Lu, R. M. Ross, M. Stagnaro, J. Zhang *et al.*, "Understanding and reducing online misinformation across 16 countries on six continents," 2022.

[4] P. Atanasova, "Generating fact checking explanations research output: Chapter in book/report/conference proceeding¿ article in proceedings¿ research¿ peer-review," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics*, 2020, pp. 7352–7364.

[5] M. Basol, J. Roozenbeek, and S. van der Linden, "Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news," *Journal of cognition*, vol. 3, no. 1, 2020.

[6] Y. Benkler, R. Faris, and H. Roberts, *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018.

[7] H. Berghel, "Malice domestic: The cambridge analytica dystopia," *Computer*, vol. 51, no. 5, pp. 84–89, 2018.

[8] D. M. Beskow, S. Kumar, and K. M. Carley, "The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning," *Information Processing & Management*, vol. 57, no. 2, p. 102170, 2020.

[9] S. Bhatt, S. Joglekar, S. Bano, and N. Sastry, "Illuminating an ecosystem of partisan websites," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 545–554. [Online]. Available: https://doi.org/10.1145/3184558.3188725

[10] L. J. Camp, "Mental models of privacy and security," *IEEE Technology and Society Magazine*, vol. 28, no. 3, pp. 37–46, 2009.

[11] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The covid-19 social media infodemic," *Scientific Reports*, vol. 10, no. 1, p. 16598, 2020.

[12] K. Clayton, S. Blair, J. A. Busam, S. Forstner, J. Glance, G. Green, A. Kawata, A. Kovvuri, J. Martin, E. Morgan *et al.*, "Real solutions for fake news? measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media," *Political Behavior*, pp. 1–23, 2019.

[13] J. Colliander, ""this is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media," *Computers in Human Behavior*, vol. 97, pp. 202–215, 2019.

[14] P. Cudré-Mauroux, "Leveraging social context for fake neoods detection: Technical perspective," *Commun. ACM*, vol. 65, no. 4, p. 123, mar 2022. [Online]. Available: https://doi.org/10.1145/3517213

[15] R. d'Andrade, "A folk model of the mind," *Cultural models in language and thought*, pp. 112–148, 1987.

[16] U. K. H. Ecker, S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen, "The psychological drivers of misinformation belief and its resistance to correction," *Nature Reviews Psychology*, vol. 1, no. 1, pp. 13–29, 2022.

[17] J. Edson C Tandoc, D. Lim, and R. Ling, "Diffusion of disinformation: How social media users respond to fake news and why," *Journalism*, vol. 21, no. 3, pp. 381–398, 2020. [Online]. Available: https://doi.org/10.1177/1464884919868325

[18] Z. Epstein, N. Foppiani, S. Hilgard, S. Sharma, E. Glassman, and D. Rand, "Do explanations increase the effectiveness of ai-crowd generated fake news warnings?" *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, no. 1, pp. 183–193, May 2022.

[19] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi, "Tweepfake: About detecting deepfake tweets," *PLOS ONE*, vol. 16, no. 5, pp. 1–16, 05 2021.

[20] L. K. Fazio, N. M. Brashier, B. K. Payne, and E. J. Marsh, "Knowledge does not protect against illusory truth." *Journal of Experimental Psychology: General*, vol. 144, no. 5, p. 993, 2015.

[21] E. Ferrara, *Bots, Elections, and Social Media: A Brief Overview*. Cham: Springer International Publishing, 2020, pp. 95–114. [Online]. Available: https://doi.org/10.1007/978-3-030-42699-6_6

[22] M. Gambini, T. Fagni, F. Falchi, and M. Tesconi, "On pushing deepfake tweet detection capabilities to the limits," in *14th ACM Web Science Conference 2022*, ser. WebSci '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 154–163. [Online]. Available: https://doi.org/10.1145/3501247.3531560

[23] L. Grace and B. Hone, "Factitious: Large scale computer game to fight fake news and improve news literacy," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–8. [Online]. Available: https://doi.org/10.1145/3290607.3299046

[24] M. Groh, Z. Epstein, C. Firestone, and R. Picard, "Deepfake detection by human crowds, machines, and machine-informed crowds," *Proceedings of the National Academy of Sciences*, vol. 119, no. 1, p. e2110013119, 2022.

[25] Z. Guo, M. Schlichtkrull, and A. Vlachos, "A Survey on Automated

Fact-Checking," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 178–206, 02 2022.

[26] A. Hounsel, J. Holland, B. Kaiser, K. Borgolte, N. Feamster, and J. Mayer, "Identifying disinformation websites using infrastructure features," in *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*. USENIX Association, Aug. 2020. [Online]. Available: https://www.usenix.org/conference/foci20/presentation/hounsel

[27] J. Im, E. Chandrasekharan, J. Sargent, P. Lighthammer, T. Denby, A. Bhargava, L. Hemphill, D. Jurgens, and E. Gilbert, "Still out there: Modeling and identifying russian troll accounts on twitter," in *12th ACM Conference on Web Science*, ser. WebSci '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–10. [Online]. Available: https://doi.org/10.1145/3394231.3397889

[28] P. Jachim, F. Sharevski, and E. Pieroni, "TrollHunter2020: Real-time Detection of Trolling Narratives on Twitter During the 2020 US Elections," in *International Workshop on Security and Privacy Analytics 2021*, ser. IWSPA '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–11, https://doi.org/10.1145/3445970.3451158.

[29] P. Jachim, F. Sharevski, and P. Treebridge, "Trollhunter [evader]: Automated detection [evasion] of twitter trolls during the covid-19 pandemic," in *New Security Paradigms Workshop 2020*, ser. NSPW '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 59–75.

[30] Y. Jeon, B. Kim, A. Xiong, D. LEE, and K. Han, "Chamberbreaker: Mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, oct 2021. [Online]. Available: https://doi.org/10.1145/3479859

[31] M. Johnson, S. Egelman, and S. M. Bellovin, "Facebook and Privacy: It's Complicated," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2335356.2335369

[32] J. Kirchner and C. Reuter, "Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW2, oct 2020.

[33] Z. Kou, L. Shang, Y. Zhang, and D. Wang, "Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection," vol. 6, no. GROUP, jan 2022. [Online]. Available: https://doi.org/10.1145/3492855

[34] S. Lewandowsky, J. Cook, U. Ecker, D. Albarracin, M. Amazeen, P. Kendou, D. Lombardi, E. Newman, G. Pennycook, E. Porter *et al.*, *The Debunking Handbook 2020*, 2020.

[35] S. Lewandowsky and S. van der Linden, "Countering misinformation and fake news through inoculation and prebunking," *European Review of Social Psychology*, vol. 32, no. 2, pp. 348–384, 2021.

[36] G. Lima, J. Han, and M. Cha, "Others are to blame: Whom people consider responsible for online misinformation," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW1, apr 2022.

[37] R. Maertens, J. Roozenbeek, M. Basol, and S. van der Linden, "Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments." *Journal of Experimental Psychology: Applied*, vol. 27, no. 1, p. 1, 2021.

[38] H. Matatov, M. Naaman, and O. Amir, "Stop the [image] steal: The role and dynamics of visual content in the 2020 u.s. election misinformation campaign," vol. 6, no. CSCW2, nov 2022. [Online]. Available: https://doi.org/10.1145/3555599

[39] M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 13, pp. 2078–2091, 2007.

[40] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," in *Proceedings of the 28th ACM International Conference on Multimedia*, ser. MM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 2823–2832. [Online]. Available: https://doi.org/10.1145/3394171.3413570

[41] B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. 2, pp. 303–330, 2010.

[42] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 6086–6093. [Online]. Available: https://aclanthology.org/2020.lrec-1.747

[43] I. Paraschivoiu, J. Buchner, R. Praxmarer, and T. Layer-Wagner, "Escape the fake: Development and evaluation of an augmented reality escape room game for fighting fake news," in *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 320–325. [Online]. Available: https://doi.org/10.1145/3450337.3483454

[44] D. Parekh, D. Margolin, and D. Ruths, "Comparing audience appreciation to fact-checking across political communities on reddit," in *12th ACM Conference on Web Science*, ser. WebSci '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 144–154. [Online]. Available: https://doi.org/10.1145/3394231.3397904

[45] G. Pennycook, T. D. Cannon, and D. G. Rand, "Prior exposure increases perceived accuracy of fake news." *Journal of experimental psychology: general*, vol. 147, no. 12, p. 1865, 2018.

[46] G. Pennycook, Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand, "Shifting attention to accuracy can reduce misinformation online," *Nature*, vol. 592, no. 7855, pp. 590–595, 2021.

[47] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," *Psychological Science*, vol. 31, no. 7, pp. 770–780, 2020.

[48] G. Pennycook and D. G. Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, vol. 188, pp. 39–50, 2019, the Cognitive Science of Political Thought. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001002771830163X

[49] ——, "The psychology of fake news," *Trends in Cognitive Sciences*, vol. 25, no. 5, pp. 388–402, 2021.

[50] F. Pierri, C. Piccardi, and S. Ceri, "A multi-layer approach to disinformation detection in us and italian news spreading on twitter," *EPJ Data Science*, vol. 9, no. 1, p. 35, 2020.

[51] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *arXiv preprint arXiv:1702.05638*, 2017.

[52] M. pui Sally Chan, C. R. Jones, K. H. Jamieson, and D. Albarracín, "Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation," *Psychological Science*, vol. 28, no. 11, pp. 1531–1546, 2017.

[53] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 1326–1343.

[54] M. D. Rich *et al.*, *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Rand Corporation, 2018.

[55] M. H. Saeed, S. Ali, J. Blackburn, E. D. Cristofaro, S. Zannettou, and G. Stringhini, "Trollmagnifier: Detecting state-sponsored troll accounts on reddit," in *2022 IEEE Symposium on Security and Privacy (SP)*, 2022, pp. 2161–2175.

[56] B. Schiller, J. Daxenberger, and I. Gurevych, "Stance detection benchmark: How robust is your stance detection?" *KI - Künstliche Intelligenz*, vol. 35, no. 3, pp. 329–341, 2021.

[57] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The Limitations of Stylometry for Detecting Machine-Generated Fake News," *Computational Linguistics*, vol. 46, no. 2, pp. 499–510, 06 2020.

[58] E. Shaabani, A. Sadeghi Mobarakeh, H. Alvari, and P. Shakarian, "An end-to-end framework to identify pathogenic social media accounts on twitter," in *2019 2nd International Conference on Data Intelligence and Security (ICDIS)*, 2019, pp. 128–135.

[59] F. Sharevski, R. Alsaadi, P. Jachim, and E. Pieroni, "Misinformation warnings: Twitter's soft moderation effects on covid-19 vaccine belief

echoes," *Computers & Security*, vol. 114, p. 102577, 2022, https://doi.org/10.1016/j.cose.2021.102577.

[60] F. Sharevski, A. Devine, P. Jachim, and E. Pieroni, ""Gettr-ing" User Insights from the Social Network Gettr," 2022, https://truthandtrustonline.com/wp-content/uploads/2022/10/TTO_2022_proceedings.pdf.

[61] ——, "Meaningful context, a red flag, or both? preferences for enhanced misinformation warnings among us twitter users," in *Proceedings of the 2022 European Symposium on Usable Security*, ser. EuroUSEC '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 189–201, https://doi.org/10.1145/3549015.3555671.

[62] F. Sharevski, P. Jachim, E. Pieroni, and N. Jachim, "Voxpop: An experimental social media platform for calibrated (mis)information discourse," in *New Security Paradigms Workshop*, ser. NSPW '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 88–107. [Online]. Available: https://doi.org/10.1145/3498891.3498893

[63] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, ser. WSDM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 312–320. [Online]. Available: https://doi.org/10.1145/3289600.3290994

[64] M. Sleeper, J. Cranshaw, P. G. Kelley, B. Ur, A. Acquisti, L. F. Cranor, and N. Sadeh, ""I Read My Twitter the next Morning and Was Astonished": A Conversational Perspective on Twitter Regrets," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 3277–3286. [Online]. Available: https://doi.org/10.1145/2470654.2466448

[65] L. Soetekouw and S. Angelopoulos, "Digital resilience through training protocols: Learning to identify fake news on social media," *Information Systems Frontiers*, 2022.

[66] K. Starbird, "Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, no. 1, pp. 230–239, May 2017.

[67] K. Starbird, A. Arif, and T. Wilson, "Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, nov 2019. [Online]. Available: https://doi.org/10.1145/3359229

[68] L. G. Stewart, A. Arif, and K. Starbird, "Examining trolls and polarization with a retweet network," in *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web. 2018.*, 2018.

[69] G. Stringhini, "Computational methods to understand and mitigate online aggression." USENIX Association, Feb. 2021.

[70] B. Swire-Thompson, J. DeGutis, and D. Lazer, "Searching for the backfire effect: Measurement and design considerations," *Journal of Applied Research in Memory and Cognition*, vol. 9, no. 3, pp. 286–299, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211368120300516

[71] W. Theisen, J. Brogan, P. B. Thomas, D. Moreira, P. Phoa, T. Weninger, and W. Scheirer, "Automatic discovery of political meme genres with diverse appearances," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, no. 1, pp. 714–726, May 2021.

[72] J. Tidwell, *Designing Interfaces: Patterns for Effective Interaction Design*. O'Reilly Media, 2010.

[73] M. Tully, L. Bode, and E. K. Vraga, "Mobilizing users: Does exposure to misinformation and its correction affect users' responses to a health misinformation post?" *Social Media + Society*, vol. 6, no. 4, p. 2056305120978377, 2020.

[74] H. Tumber and S. Waisbord, *The Routledge companion to media disinformation and populism*. Routledge, 2021.

[75] B. Ur, F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin, and L. F. Cranor, ""I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Jul. 2015, pp. 123–140.

[76] P. van Schaik, J. Jansen, J. Onibokun, J. Camp, and P. Kusev, "Security and privacy in online social networking: Risk perceptions and precautionary behaviour," *Computers in Human Behavior*, vol. 78, pp. 283–297, 2018.

[77] G. Verma, A. Bhardwaj, T. Aledavood, M. De Choudhury, and S. Kumar, "Examining the impact of sharing covid-19 misinformation online on mental health," *Scientific Reports*, vol. 12, no. 1, p. 8045, 2022.

[78] M. Volkamer and K. Renaud, *Mental Models – General Introduction and Review of Their Application to Human-Centred Security*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 255–280.

[79] R. Wash, "Folk Models of Home Computer Security," in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, ser. SOUPS '10. New York, NY, USA: Association for Computing Machinery, 2010. [Online]. Available: https://doi.org/10.1145/1837110.1837125

[80] R. Wash and E. Rader, "Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, Jul. 2015, pp. 309–325. [Online]. Available: https://www.usenix.org/conference/soups2015/proceedings/presentation/wash

[81] M. Wigell, "Hybrid interference as a wedge strategy: A theory of external interference in liberal democracy," *International Affairs*, vol. 95, no. 2, pp. 255–275, 5/26/2022 2019.

[82] T. Wilson, K. Zhou, and K. Starbird, "Assembling strategic narratives: Information operations as collaborative work within an online community," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, nov 2018.

[83] T. Wood and E. Porter, "The elusive backfire effect: Mass attitudes' steadfast factual adherence," *Political Behavior*, vol. 41, no. 1, pp. 135–163, 2019. [Online]. Available: https://doi.org/10.1007/s11109-018-9443-y

[84] L. Wu, F. Morstatter, K. M. Carley, and H. Liu, "Misinformation in social media: Definition, manipulation, and detection," *SIGKDD Explor. Newsl.*, vol. 21, no. 2, p. 80–90, nov 2019. [Online]. Available: https://doi.org/10.1145/3373464.3373475

[85] S. Zannettou, T. Caulfield, J. Blackburn, E. De Cristofaro, M. Sirivianos, G. Stringhini, and G. Suarez-Tangil, "On the origins of memes by means of fringe web communities," in *Proceedings of the Internet Measurement Conference 2018*, ser. IMC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 188–202. [Online]. Available: https://doi.org/10.1145/3278532.3278550

[86] S. Zannettou, T. Caulfield, B. Bradlyn, E. De Cristofaro, G. Stringhini, and J. Blackburn, "Characterizing the use of images in state-sponsored information warfare operations by russian trolls on twitter," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, no. 1, pp. 774–785, May 2020. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/7342

[87] S. Zannettou, T. Caulfield, W. Setzer, M. Sirivianos, G. Stringhini, and J. Blackburn, "Who let the trolls out? towards understanding state-sponsored trolls," in *Proceedings of the 10th ACM Conference on Web Science*, ser. WebSci '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 353–362.

[88] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, "The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans," *J. Data and Information Quality*, vol. 11, no. 3, may 2019. [Online]. Available: https://doi.org/10.1145/3309699

[89] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," in *Proceedings of the Web Conference 2021*, ser. WWW '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 3465–3476. [Online]. Available: https://doi.org/10.1145/3442381.3450004

APPENDIX

A. Study Questionnaire

Introduction:

0. Misinformation on social media is an umbrella term that includes all false or inaccurate information that is spread on social network platforms, such as: disinformation, fake news, rumors, conspiracy theories, hoaxes, trolling, urban legends, and spam [84].

Exposure and Preconceptions:

1. What is misinformation, in your personal view? **[Open Ended]**
2. Could you please specify all the platforms where you have encountered misinformation and, if possible, provide some examples **[Twitter, Facebook, Reddit, Gab, Gettr, Parler, Rumbler, Truth Social]**
3. What was your initial response to some of the misinformation examples you have provided? Please elaborate. **[Open Ended]**
4. Where does misinformation on social media come from, in your opinion? **[Open Ended]**
5. What kind of function does the misinformation on social media serve, in your opinion? **[Open Ended]**
6. Who benefits from misinformation on social media, in your opinion? **[Open Ended]**

Engagement Strategies:

7. How do you suspect/know a certain social media post is misinformation? Please elaborate. **[Open Ended]**
8. What is your strategy for dealing with misinformation posts on social media? Please elaborate. **[Open Ended]**
9. Are there occasions where you have, or are inclined to, comment/reply to a misinformation post? If so, what did you or would you say in your comment/reply? Please elaborate **[Open Ended]**
10. Are there occasions where you have, or are inclined to, use any engagement features (e.g. like, retweet/repost, share, follow) when encountering a misinformation post? If so, in what circumstances? Please elaborate **[Open Ended]**
11. Are there occasions where you have, or are inclined to, use any action features (e.g. block, mute, report, unfollow) when encountering a misinformation post? If so, in what circumstances? Please elaborate **[Open Ended]**
12. Are there occasions where you have or are inclined to talk about a particular misinformation post outside social media? If so, in what circumstances? Please elaborate **[Open Ended]**
13. Are there occasions where you have or are inclined to engage with a misinformation post using counter-argumentation? If so, in what circumstances? Please elaborate **[Open Ended]**
14. Are there occasions where you have or are inclined to engage with a misinformation post using humor, sarcasm, mocking, or taunting? If so, in what circumstances? Please elaborate **[Open Ended]**

Follow-up:

15. We plan a voluntary follow-up with anyone if they are interested in expanding on their ways of dealing with misinformation. If you like to do so, please provide your email contact. This won't affect your previous participation and compensation and the data will be incorporated in a way that cannot be linked back to you.

B. Codebook

Political (Counter)Argumentation

TABLE II: **Folk Model Definition**

Definition	Any information that has faithfulness to *selective* facts relative to political and ideological contexts, created and disseminated with agenda-setting or argument-winning intentionality
Inclusion Criteria	Any response about misinformation that explicitly points to a political involvement in use of information containing selective facts and ideologically-biased argumentation or counter-argumentation
Example Response	"*Made-up stories of politicians or government policies with fake news being shared as facts*" [**P81FA40**]

TABLE III: **Origins of Misinformation**

Definition	Spreaders of misinformation that take and/or represent, defend, and argue for one political ideology against others
Inclusion Criteria	Any response indicating political polarization in originating and spreading alternative narratives
Example Response	"*Narrative from the 'other side' that paint the political opponents incorrect and bad*" [**P193FL40**]

TABLE IV: **Misinformation Purpose**

Definition	Misinformation weaponized for political (counter)argumentation.
Inclusion Criteria	Any response explicitly indicating misinformation's purpose to be for political means
Example Response	"*Misinformation serves to foster division between ideologies, prop up 'straw man' arguments, and advocate for particular legal and judicial outcomes*" [**P32ML50**]

Out-of-context Narratives

TABLE V: **Folk Model Definition**

Definition	Any information that has *questionable* faithfulness to known facts due to selection of improbable alternative contexts, created and disseminated with speculative intentions
Inclusion Criteria	Any response that explicitly refers to manipulative misconstruction of facts with speculative intentions
Example Response	"*Information on Twitter about studies that have been misrepresented or taken out of context, Facebook shared "news stories" from questionable perspectives, and Reddit contains commentary on news stories that are misconstrued*" [**P188MR40**]

TABLE VI: **Origins of Misinformation**

Definition	Spreaders of misinformation that manipulate facts, selectively omit them, misinterpret, or use them in a speculative manner
Inclusion Criteria	Any response indicating manipulative speculation and interplay with facts as an origin for misinformation
Example Response	"*Twisting what's actually a personal opinion into one's subjective idea of a fact*" [**P200FM30**]

TABLE VII: **Misinformation Purpose**

Definition	Misinformation weaponized for sowing discord among people.
Inclusion Criteria	Any response indicating the divisive and polarising purpose of misinformation outside of the political arena
Example Response	"*To deceive and deflect there own responsibility for many of the issues we face*" [**P186FL50**]

Inherently Fallacious Information

TABLE VIII: **Folk Model Definition**

Definition	Any information *unfaithful* to to known facts, regardless of contexts or intentions
Inclusion Criteria	Any response that explicitly indicates misinformation as false or inaccurate information and does not assign political or manipulative motif to it
Example Response	"*Blatant falsehoods with no touch to the reality*" [**P57FL50**]

TABLE IX: **Origins of Misinformation**

Definition	Misinformation solely comprised of falsehoods, fabrications, or inaccuracies
Inclusion Criteria	Any response indicating misinformation is information that does not include any known facts
Example Response	"*False information regarding COVID-19.*" [**P142FM50**]

TABLE X: **Misinformation Purpose**

Definition	Misinformation weaponized for polluting the information-sharing systems
Inclusion Criteria	Any response indicating use of misinformation for the purpose of polarization, division, and discord
Example Response	"*falsehoods that usually stir up controversy or sow dissension amongst the masses*" [**P90FL60**]

External Propaganda

TABLE XI: **Folk Model Definition**

Folk Model Definition	Any information with a *fluctuating* faithfulness to known facts relative to shifting contexts, created and disseminated with a propagandistic intentions
Inclusion Criteria	Any response that explicitly points to propaganda on social media
Example Response	"*Russian bots trying to spew pro Russian propaganda about why they attacked Ukraine*" [**P120MM40**]

TABLE XII: **Origins of Misinformation**

Definition	Spreaders of disinformation and information operations on behalf of other nation state governments
Inclusion Criteria	Any response indicating misinformation intentionally disseminated falsehoods for the purpose to mislead
Example Response	"*I believe a lot of it comes from hostile nation states like Russia, China, and Iran. It is then taken up by groups open to the message and then spreads like a disease*" [**P60MR40**]

Entertainment

TABLE XIII: **Misinformation Purpose**

Definition	Misinformation in any form used for information operations and disinformation campaigns
Inclusion Criteria	Any response explicitly seeing misinformation's purpose as information operations of disinformation campaigns
Example Response	"*Propaganda permeated by bots from nation-states to cause rifts, hate, and sow discord in their enemy's backyard*" [**P112ML40+**]

TABLE XIV: **Folk Model Definition**

Folk Model Definition	Any information with a *tangential* faithfulness to known facts relative to humorous or sarcastic contexts, usually created and disseminated with entertaining intentions
Inclusion Criteria	Any response that explicitly points to misinformation as entertainment, humor, or jokes
Example Response	"*People making jokes, for example is Elon Musk saying he is going to put cocaine in cola*" [**P233MM40**]

TABLE XV: **Origins of Misinformation**

Definition	Misinformation solely used for entertainment and excluding explicit notions to hate speech, offensive language, and derogatory multimedia content
Inclusion Criteria	Any response explicitly referring to misinformation-as-entertainment
Example Response	"*People making jokes and other people believing them as real*" [**P233FM40**]

TABLE XVI: **Misinformation Purpose**

Definition	Misinformation in any form used for entertainment purposes
Inclusion Criteria	Any response explicitly seeing misinformation's purpose exclusively for entertaining social media users
Example Response	"*Entertainment purposes. If it's really bizarre people will read it*" [**P133FM61+**]