emerald PUBLISHING

Internet Re

Online Silence: why do not people challenge others when posting misinformation?

Journal:	*Internet Research*
Manuscript ID	INTR-06-2022-0407.R3
Manuscript Type:	Research Paper
Keywords:	misinformation, challenging misinformation, online silence, social media

SCHOLARONE™
Manuscripts

Online Silence: Why Do Not People Challenge Others When Posting Misinformation?

Abstract

Purpose: There is a scarcity of research studies on why people remain inactive when encountering and recognising misinformation online. The main aim of this paper is to provide a groundwork for future research into why users do not challenge misinformation on digital platforms by generating hypotheses through a synthesis of pertinent literature, including organisational behaviour, communication, human-computer interaction (HCI), psychology, and education.

Design/methodology/approach: Given the lack of directly related literature, this paper synthesised findings from relevant fields where the findings might be relevant, as the tendency to withhold opinions or feedback is a well-documented practice in offline interaction.

Findings: Following our analysis of relevant literature, the potential reasons for online silence towards misinformation can be divided into six categories: self-oriented, relationship-oriented, others-oriented, content-oriented, individual characteristics, and technical factors.

Originality: Although corrections coming from peers can effectively combat misinformation, several studies showed that people in cyberspace do not take such action. To the best of our knowledge, there has been scarce and virtually non-existent research investigating why people refrain from challenging others who post misinformation online. Thus, this paper attempts to address this gap and identify reasons in adjacent domains. The reasons provide a starting point for researching interventions to reduce reluctance and abstinence regarding the challenge of misinformation. Our findings can be beneficial beyond the area of challenging misinformation and are extensible to other types of content and communication that people are hesitant to discuss and challenge, such as online injustice, prejudice, and hate speech.

Keywords—misinformation, challenging misinformation, digital platforms, online silence

I. INTRODUCTION

Many users on digital platforms may not see themselves as key players in mitigating misinformation, but many may contribute to its dissemination through their silence. One of the main challenges in combatting misinformation is that when users notice it online, most ambivalently ignore it and do not respond to others to challenge it (Chadwick and Vaccari 2019; Vicol 2020). We call this behaviour "Online Silence". One might argue that, in not challenging such misinformation, users are complicit in its spread by being silent. In this context, silence is construed as a type of self-censorship, which refers to the choice not to speak out despite the belief that something should be said (Pinder and Harlos 2001). The way social media limits the exposure of diverse opinions and promotes a common narrative with like-minded people, known as an "echo chamber" (Cinelli *et al.* 2021), is also argued to amplify users' cognitive biases, such as confirmation bias, which refers to the tendency to search or remember information that confirms or supports one's previous opinions (Westerwick *et al*. 2017; Zhao *et al*. 2020). Thus, those who agree with the mindset of the false information, knowing it is

indeed false, and see no harm in spreading it, may not only remain silent but also help to disseminate it by showing their support for the content through “liking” or commenting positively on it. Studying this group of users and their motives is beyond our scope as we only concentrate on self-censorship, which refers to remaining silent and suppressing the will to respond when something should be said.

With the widespread use of the internet and Social Networking Sites (SNS) for information seeking and users becoming content creators and broadcasters on such sites, information starts to spread widely even before its accuracy can be verified. Although many terms such as disinformation, fake news, and rumours are used interchangeably to refer to incorrect or misleading information, throughout this paper, the term “misinformation” will be used as an umbrella term, as it refers widely to any type of false information, regardless of the intention or format.

On SNS, fake news spreads six times faster than true news (Vosoughi *et al.* 2018) and misinformation is engaged with more than factual posts (Edelson *et al.* 2021). A recent study also showed that using words related to conspiracy theories in a tweet increases its chances of being shared (Visentin *et al*. 2021). In such an environment, it becomes even more important to combat the spread of misinformation. Even though prior research showed that approximately 5% of people's news consumption is comprised of misinformation (Acerbi *et al.* 2022), the effects of information operations should not be underestimated as the spectrum of ramifications and potential problems is quite broad. This spectrum ranges from affecting consumers’ attitudes toward brand image (Visentin *et al*. 2019; Borges-Tiago *et al*. 2020) and consumers’ purchase intentions (Mishra and Samu 2021) to altering people’s attitudes toward issues such as climate change (Lutzke *et al*. 2019) or voting behaviour (Cantarella *et al*. 2023). Social media companies took proactive measures to combat misinformation after the 2016 United States election prompted concerns over misinformation online (Pourghomi *et al.* 2017). However, much as companies and scholars have raised concerns about the issue, the spread and severe impact of misinformation are still being seen in critical domains, e.g. public health information related to the COVID-19 pandemic (Cinelli *et al*. 2020; Galvão 2021) and political and humanitarian domains, e.g. the Ukraine-Russia war (Park *et al.* 2022). This points to a need to supplement these measures with new approaches and techniques employing individuals as not only reporters but also activists in challenging misinformation.

"Social corrections," which refer to correction attempts made by social sources where social contacts are usually a primary source of information, have been identified as a possible intervention (Bode and Vraga 2018; Walter and Murphy 2018; Bode and Vraga 2021b; Walter *et al*. 2021). It reduces the spread of misinformation by influencing other users who observe the corrections (Vraga and Bode 2018). It is even more effective when users provide credible sources to refute the information provided (Vraga and Bode 2017). Therefore, it is important that users who post misinformation are informed about this when possible. However, although evidence shows that users’ corrections are as effective as algorithmic corrections (Bode and Vraga 2018), people can be hesitant to take any action to correct misinformation (Chadwick and Vaccari 2019; Tandoc *et al*. 2020; Tully *et al*. 2020).

Research that studied social corrections as an intervention strategy to combat misinformation lacked the investigation of why people refrain from correcting misinformation or even do not feel the need to make that correction in the first place. Existing research has mainly focused

on strategies to motivate people to challenge misinformation rather than the barriers that prevent users from doing so. Evidence suggests that exposure to norms (i.e., whether individuals believe others typically correct misinformation) (Koo *et al.* 2021) or their loved ones would want them to correct (Xiao 2022), people's perception of the severity of the influence of misinformation on others (Sun *et al.* 2021) and social identity threat (i.e., when misinformation threatens a person's social status in a group) (Cohen *et al.* 2020) seem to be effective in motivating people to correct misinformation. Factors such as interpersonal relationships with the poster, where, for example, participants were more likely to correct friends or family members (Tandoc *et al.* 2020), and the format of the content, where, for example, individuals were less likely to correct content presented in meme format (Lyons 2017) also seem to affect users' intention to challenge others when posting misinformation. In addition, Bode and Vraga (2021a) showed that social factors such as age, education, and reliance on mainstream news and social media also affect willingness to challenge misinformation. A qualitative study that interviewed participants in Vietnam revealed that the degree of closeness with the sharer, the age of the sharer, and also whether the environment in which correction takes place is public or private, impact the decision to correct misinformation (Rohman 2021).

Previous studies exploring the barriers that prevent or discourage people from challenging misinformation were limited in scope. Following a series of interviews in the United States with physicians and nurses, the obstacles that they face when they correct health misinformation on social media were identified in three categories: intrapersonal (e.g., the lack of time and the perception of limited positive outcomes), interpersonal (e.g., fear of being harassed and bullied), and institutional (e.g., a lack of institutional support and social media training) (Bautista *et al.* 2021). Another recent study based on interviews with 102 people in the U.K regarding COVID vaccines also found that the social norm of conflict avoidance affects people's responses to vaccine misinformation (Chadwick *et al.* 2022).

While prior studies present a broad range of factors that motivate people to challenge misinformation, it is also essential to identify the barriers that prevent them from doing so and, consequently, develop interventions to eliminate those barriers as a preliminary step. Offering motivational strategies does not necessarily eliminate people's barriers to challenging misinformation. For example, disclaimers from social media platforms (Colliander 2019) or source ratings, as suggested in (Kim *et al*. 2019), do not necessarily serve to overcome barriers such as fear of receiving a hostile response from the one who shared misinformation or fear of harming the relationship with them.

The findings of this paper pave the way for future research to understand the barriers that prevent users from remaining silent when they see misinformation. Furthermore, exploring these factors that have been overlooked previously might be a starting point for enhancing the current design of digital platforms. The Persuasive Systems Design (PSD) model (Oinas-Kukkonen and Harjumaa 2009) has been used in various domains to promote behaviour change, such as health (Orji and Moffatt 2018), fitness (Oyibo and Vassileva 2019) and e-learning (Widyasari *et al.* 2019). It provides principles and techniques to design socio-technical solutions that can influence attitudes and behaviour by prompting cognitive, behavioural, psycho-social, and other psychological processes. PSD strategies are extensive and based on well-established theories of behaviour change. As the model has already been used to change online behaviour, e.g. gaming (Alrobai *et al.* 2016; Adib *et al.* 2021) and cyber security (Misra *et al.* 2017), it can also be a helpful reference model in designing for motivating users

to challenge misinformation. The purpose of this paper is twofold. The first is to identify reasons for not challenging misinformation and to pave the way for future research on the topic, and the second is to propose persuasive socio-technical interventions to motivate users to speak up when they encounter misinformation on SNS.

In this research, the term "challenging" rather than "correcting" is used for two reasons. First, "correcting" is an absolute statement that presumes the person doing the correction is accurate; however, the person who intends to correct may not actually be correct. Second, this paper does address not only corrections but also disagreements or disputes over the content. In other words, the term "challenging" serves a broader function than simply correcting.

Taken together, this work is a starting point of a broader study that aims to contribute to altering the trend of seeing misinformation but not questioning it. This research does not intend to be a systematic review, scoping review, or narrative review as to our knowledge, there is no literature primarily focused on this topic. This paper aims to synthesise pertinent literature on related topics and provide potential reasons regarding users' silence towards misinformation online. This paper is organised as follows. Section 2 presents a literature review and an overview of challenging misinformation. Section 3 provides insights into possible reasons people are reluctant to challenge misinformation. We conclude and present possible solutions and future work directions in Section 4.

II. Background and Motivation

As many as 53% of U.S. adults obtain news from social media (Shearer and Mitchell 2021). Besides being a fertile ground for misinformation, social media also offers opportunities to mitigate the problem (Katie Elson 2018; Djordjevic 2020). In addition to algorithmic approaches or machine learning-based solutions, individuals' active participation in conversations to challenge misinformation can help reduce misinformation (Bode and Vraga 2018; Margolin *et al.* 2018).

Data from several studies suggests that correcting misinformation is not common on social media. Almost 80% of social media users in the U.K. have not told anyone who shared false news on social media that the news they shared was false or exaggerated (Chadwick and Vaccari 2019). Another U.K. survey found that although 58% of respondents reported having encountered content that they thought was false, only 21% said they did something to correct it (Vicol 2020). Recent research in the U.S. regarding the correction of misinformation about COVID-19 on social media revealed similar findings. Among 56.6% of those reporting that they saw misinformation, only 35.1% said they corrected someone (Bode and Vraga 2021c). Similarly, in Singapore, 73% of social media users dismiss fake news posts on social media without taking further action (Tandoc *et al.* 2020). Some studies have documented that actively interacting with corrections when exposed to unconfirmed claims is rare (Zollo *et al*. 2017), SNS users are not consistently motivated to correct misinformation publicly (Cohen *et al.* 2020), and explicit corrections of other users are rare in the online environment (Arif *et al.* 2017).

Reporting misinformation is one of the techniques provided by social media, enabling individuals to mark a post as false anonymously. However, much as reporting helps diminish

the problem, it requires several steps and does not allow users to express their opinions; thereby, it does not help to generate a constructive and meaningful dialogue. Such dialogue may also have the extra benefit of altering the beliefs and enhancing the critical literacy of the social sources posting or sharing misinformation and their audience, which can, in turn, foster long-term behaviour change. Active engagement with false posts to challenge them by deliberation, argumentation, or questioning is crucial to decreasing misinformation dissemination and cultivating a diverse environment as it increases distinct ideas. Individuals modify their beliefs about misinformation after seeing another user being corrected on social media (Vraga and Bode 2017). Users are also less likely to spread rumours when they feel they could be confronted with a counterargument, criticism, or warning (Tanaka *et al.* 2013; Ozturk *et al.* 2015) and are more likely to comment critically on posts when they are exposed to critical comments from other users (Colliander 2019).

III. Potential Reasons for Avoiding Challenging Misinformation Online

The online sphere was construed as fundamentally different from offline spaces by offering an environment where individuals are less restricted from expressing ideas. Some believe that the risks of expressing opinions and sharing content online are lower than doing the same offline (Papacharissi 2002; Ho and McLeod 2008; Luarn and Hsieh 2014). However, according to social information processing theory (Walther 1992), although interpersonal development requires more time in a computer-mediated environment than face-to-face (FtF), users may develop similar levels of interpersonal relations. This theory suggests that, regardless of the medium, people are driven to form impressions and build connections, and that language or symbols in computer-mediated communication (CMC) are as important as nonverbal cues in FtF communication. Based on these findings, the constraints people have when expressing contradictory opinions or challenging people in an online environment might not differ from those in person.

In FtF communication, withholding opinions, abstaining from participating in discussions, or remaining silent even though there is an issue that needs intervention may all occur. Refraining from providing opinions occurs in various environments, such as enterprise environments where employees withhold information purposefully (Morrison and Milliken 2000; Dyne *et al.* 2003; Milliken *et al.* 2003) or in classrooms, where students keep silent during discussions (Fassinger 1995; Jaworski and Sachdev 1998; Rocca 2010).

The silence of employees in organisations is considered to be driven by several motivations and should not be taken as a sign of acceptance. According to Pinder and Harlos (2001) and Dyne *et al*. (2003), there are three types of silence based on employee motives: acquiescent, defensive, and prosocial. Acquiescent silence is passive behaviour and is motivated by a lack of desire to speak up. An employee could withhold their ideas due to the belief that speaking up will not change the situation or that they are unable to make a difference. Defensive silence is described as proactive behaviour and is motivated by the intention of protecting oneself. An employee could remain silent due to fear of the consequences of expressing ideas, such as getting fired or demoted. Finally, prosocial silence is defined as withholding ideas based on positive intentions for others or the organisation. An employee could be reticent due to a desire to protect others from embarrassment or trouble.

Internet Research

In a qualitative study investigating why employees remain silent (Milliken *et al.* 2003), fear of being viewed or labelled negatively and damaging valued relationships were the most frequently mentioned reasons. Even in situations where silence might have devastating consequences, people might still choose not to speak up. Bienefeld and Grote (2012) revealed that although speaking up is critical for flight safety; aircrew members are reluctant to do so because of the adverse outcomes of speaking up. Their most common reason for not speaking was their desire to maintain a good relationship with the team and not lose the other crew members' acceptance and trust. In addition, captains were afraid of embarrassing first officers, and first officers were concerned that captains would view them as troublemakers if they contradicted captains. Reasons that hinder employees from speaking up align closely with the educational psychology literature investigating reasons for student participation in the classroom. Barriers to participating in class range from negative outcome expectations or evaluation apprehension to fear of appearing unintelligent or inadequate to one's peers or instructors (Fassinger 1995; Rocca 2010). Logistics such as class size, seating arrangement, mandatory participation, and the instructor's influence also affect students' participation (Rocca 2010). Taken together, these studies provide important insights into why people refrain from entering conversations, speaking up, or questioning in offline environments.

In CMC, users also choose to be silent. They refrain from discussing their ideas (Hampton *et al.* 2014), posting content about political and social issues (McClain 2021), commenting on questionable news (Stroud *et al.* 2016) and correcting misinformation (Chadwick and Vaccari 2019; Tandoc *et al.* 2020). The concept of silence or non-participation in the online environment is conceptualised as lurking or passive SNS use, which refers to passively viewing and not posting or participating in an online community (Nonnecke and Preece 2001). In trying to understand the factors affecting lurking behaviour, Amichai-Hamburger *et al*. (2016) proposed a model with three main reasons for remaining passive: individual differences (need for gratification, personality dispositions, time available, and self-efficacy), social-group processes (socialisation, type of community, social loafing, responses to delurking, and quality of the response), and technological setting (technical design flaws and the privacy and safety of the group). However, the concept of lurking or passive SNS use may not provide a comprehensive explanation of self-silencing behaviour in the online environment, as self-silencing might be due to reasons other than just passively browsing or choosing to observe over participating.

There is a lack of evidence on what prevents people from challenging online misinformation. Such active behaviour can effectively complement existing technical and socio-technical solutions, such as those based on A.I. and natural language processing (NLP) (de Oliveira *et al.* 2021), the analysis of the profile of media sources (Nakov 2020), and crowdsourcing that enables the reporting of misinformation (Kim *et al.* 2018). The active role of users can also combat misinformation beyond the public online forums and cover closed platforms such as messaging groups, and possibly cover languages and dialects that current A.I. and NLP solutions do not cover. We first scanned the literature to identify relevant papers to our mission in identifying reasons for not challenging misinformation. We used combinations of keywords to search for articles which contained them in their title, abstract or keywords list. The search sentence we used was [factors OR reasons OR determinants OR barriers] AND [challeng* OR question* OR correct* OR counteract* OR debunk* OR refut*] AND [misinformation OR fake news OR disinformation OR rumour OR rumor OR false information]. Our search yielded only two relevant papers, even after manually searching the proceedings of known conferences in the area covering research in media, A.I. and computational linguistics. After

that, we extended our search to include papers that studied why people remain silent when they see issues, whether misinformation or opinions, in other environments such as the workplace or classroom. These reasons can provide a starting point to understanding online silence in the context of online misinformation. Our objective was not to systematically determine which and how many articles provided a relevant result but to identify a theory-informed set of reasons for online silence by drawing a parallel with adjacent behaviours. We stopped the search when reaching a degree of saturation, i.e., finding the same identified reasons when reviewing new papers. To enhance presentation, we grouped the elicited reasons, based on their similarity, into six categories: self-oriented, relationship-oriented, others-oriented, content-oriented, individual characteristics, and technical factors (see Fig. 1.).

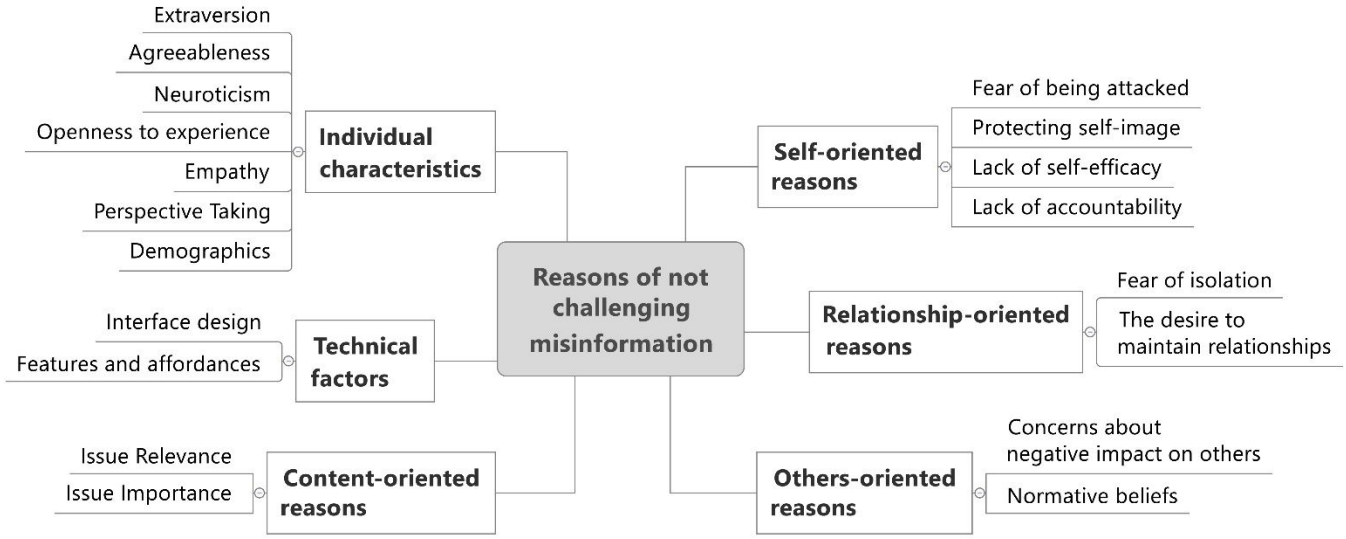

Fig.1. Potential reasons why people do not challenge misinformation

1) Self-oriented reasons

a) Fear of being attacked

Anonymity and lack of visual cues in CMC may lead to an online disinhibition effect, describing users acting less restrainedly in cyberspace than they would in real life by loosening social norms or restrictions (Suler *2004*). Internet users who are aware that cyberspace enables and provides ample opportunities for hostile communication may be reluctant to engage in challenging misinformation due to fear of being attacked or becoming the victim of cyberbullying, which is deliberate, repeated hostile behaviour to harm others using information and communication technologies (Slonje *et al.* 2013).

The negative consequences of cyberbullying are known to be intense (Barlett 2015) and include depression (Patchin and Hinduja 2006) or emotional distress (Cao *et al.* 2020). Therefore, fear of cyber aggression may thwart users from expressing their deviant opinions. For instance, college students and young adults avoid expressing their opinions regarding politics in the online environment because of online outrage (Vraga *et al.* 2015; Powers *et al.* 2019). Evidence also suggests that users exposed to cyberbullying tend to decrease or abandon their usage of SNS (Cao *et al.* 2020; Urbaniak *et al.* 2022).

The fear of being attacked might emanate from the polarisation of social media. Social media enable an environment encouraging homophily, where individuals with the same beliefs and opinions get together and become homogeneous (Cinelli *et al.* 2021). While convenient, algorithms showing users customised content based on their interests and views facilitate further polarisation. It is therefore likely that users who are aware that radicalisation and extremism are prevalent in the online environment are more prone to keeping silent. For instance, 32% of users who never or rarely share content about political or social issues cited the fear of being attacked as the reason for not posting (McClain 2021). These findings suggest that users might refrain from challenging misinformation due to fear of being attacked in the online environment.

b) Desire to protect self-image

Impression management (also known as self-presentation) is how people try to control how others see them (Leary and Kowalski 1990). According to this approach, people sometimes modify their behaviour to create positive impressions in the eyes of others by monitoring and assessing others' perceptions of themselves. People seek to manage their impressions on social media (Paliszkiewicz and Mądra-Sawicka 2016), where users are able to curate their images easily *(Weinstein 2014)*. They selectively disclose information to create a desirable, ideal, and socially acceptable image (Zhao *et al*., 2008). In order to meet the expectations of the audience, they post information that their audience will find non-offensive (Marwick and Boyd 2011) and avoid engaging in controversial topics (Sleeper *et al.* 2013).

According to impression management theory (Leary and Kowalski 1990), individuals are motivated to make a positive impression rather than act as they feel they should. In this case, it can be speculated that although individuals think they should correct misinformation (Bode and Vraga 2021c), they may remain silent owing to the risk of creating a negative impression, as conflicts, negative feedback, and political discussions on social media are not desirable (Thorson 2014; Koutamanis *et al.* 2015; Vraga *et al.* 2015).

c) *Lack* of self-efficacy

Self-efficacy theory, derived from social cognitive theory, focuses on the interconnections between behaviour, outcome expectancies and self-efficacy (Bandura 1977). According to this theory, self-efficacy refers to a person's judgement of their own ability to determine how successfully they can perform a specific behaviour. Simply put, self-efficacy is a person's belief in their own capacity to succeed.

Outcome expectancies, defined as a person's perception of the consequences of their actions, have the potential to influence self-efficacy (Bandura 1977). The theory suggests that efficacy beliefs influence outcome expectancies. More precisely, people's outcome expectancies are heavily influenced by their assessments of how well they would perform in various settings. In this case, we can speculate that individuals might not challenge misinformation due to a perceived lack of efficacy in achieving the behaviour (e.g., a perceived lack of knowledge). Indeed, when individuals feel equipped enough to express their opinions, they are more likely to speak up on a political issue regardless of what the majority thinks (Lasorsa 1991). Similarly, Tandoc *et al*. (2020) found that personal efficacy is one of the main factors affecting a user's decision to correct fake news. In sum, users may avoid challenging others due to their belief that their abilities are insufficient to succeed or that their efforts would not make any difference.

Internet Research

d) *Lack of accountability*

The bystander effect suggests that people are less likely to offer help in an emergency when other people are present because of the diffusion of responsibility (Darley and Latané 1968). Although this phenomenon is associated with emergencies in physical space, it is also examined in virtual environments (Fischer *et al.* 2011), such as participation in conversations in an online learning environment or cyberbullying (Hudson and Bruckman 2004; You and Lee 2019). It was shown that one of the reasons for not intervening in cyberbullying or not participating in conversations is that the participants delegate the responsibility for intervention to other bystanders.

In SNS, misinformation can be seen by many people, which might lead to a diffusion of responsibility in which people do not feel accountable for not correcting misinformation. They might regard their responsibility as lower because they think others might be more accountable for correcting misinformation.

2) Relationship-oriented reasons

a) Fear of isolation

Asch (Asch 1956) demonstrated empirically that individuals adjust their behaviour in order to fit in with the group. Relying on the Asch conformity experiment, Noelle-Neumann (1974) introduced the Spiral of Silence theory, which proposes that people gauge the public opinion climate and, if they perceive that their opinion is in the minority, they are more likely to hold back their opinion, while if they think their opinion is in the majority, they tend to speak out confidently. One of the main reasons for conforming is the fear of isolation. Noelle-Neumann (1974) argues that because of our social nature, we are afraid of being isolated from our peers and losing their respect. In order to avoid disapproval or social sanctions, people constantly monitor their environment and decide whether to express their opinions.

Being isolated or ostracised (ignored or excluded by others) is painful as it triggers several physiological, affective, cognitive, and behavioural responses (Williams and Nida 2011). Therefore, users on social networking sites conform, comply, or obey so they are not excluded (Williams *et al.* 2000). One of the reasons they do not correct others might be the fear of being isolated as they want to fit in the group.

b) The desire to maintain relationships

Maintaining social ties plays a pivotal role in psychological well-being (Kawachi and Berkman 2001). The need to belong is one of the fundamental needs (Williams and Sommer 1997) and is linked to psychological and physical well-being (Baumeister and Leary 1995). The pursuit of belonging also exists in cyberspace (Williams *et al.* 2000) and on SNS such as Facebook (Covert and Stefanone 2018), Instagram and Twitter (Hayes *et al*. 2018). In SNS, users interact with a large number of friends. “Friends” on SNS encompass both strong and weak ties and include friends, family, neighbours, colleagues, or romantic partners, but also acquaintances or consequential strangers (Fingerman 2009), all of whom are sources of social capital (Antheunis *et al.* 2015). One of the main motivations for using SNS is the desire to maintain relationships (Joinson 2008; Dunne *et al.* 2010). As a result, people might be more cautious in their interactions. Indeed, Gallrein *et al*. (2019) found that individuals tend to withhold negative interpersonal feedback as they perceive it has the potential to harm their relationships. In another study exploring why employees do not speak up about issues or

concerns, Milliken *et al.* (2003) reported that fear of damaging relationships is the second most common reason for withholding opinions.

As conflicts may pose a threat to users' sense of belonging, users may avoid challenging or confronting others on social media.

3) Others-oriented reasons

a) Concerns about the negative impact on others

Individuals might withhold their opinions or refrain from challenging others for altruistic purposes, such as fear of embarrassing or offending others. For instance, in organisations, employees remain silent because of the concern that speaking up might upset, embarrass or in some way harm other people (Milliken *et al.* 2003).

Withholding opinions because of concern for others reveals itself in the public arena. A Norwegian study exploring freedom of expression in different social conventions and norms found that citizens withheld their opinions in the public domain due to fear of offending others (Steen-Johnsen and Enjolras 2016). On social media, users also adhere to the norm of not offending others. A qualitative study showed that users hesitate to counteract misinformation due to fear of embarrassing the sharer, preferring to use private communication to minimise the risk (Rohman 2021).

b) Normative beliefs

Perceived norms are people's understanding of the prevalent set of rules regulating the behaviour that group members can enact (Lapinski and Rimal 2005). They can be divided into two categories: descriptive norms and injunctive norms. While descriptive norms explain beliefs regarding the prevalence of a behaviour, injunctive norms explain others' perceived approval of that behaviour (Rimal and Real 2003; Lapinski and Rimal 2005). Engagement in the behaviour is influenced by these two norms, the extent to which a behaviour is prevalent and approved by others (Berkowitz 2003).

Social norms play an important role as behavioural antecedents in many contexts, as well as in the context of the correction of misinformation. Koo *et al*. (2021) showed that when individuals perceive corrective actions as common, they are more motivated to correct misinformation. In their experimental study, Gimpel *et al*. (2021) found that emphasising the socially desirable action of reporting false news using an injunctive social norm increases the rate of reporting fake news. Although individuals think it is normative to correct someone (Bode and Vraga 2021c), it is not easy to engage in a dialogue to challenge the content poster due to the norms that govern social interactions. On Facebook, for example, where everyone's social contacts could see the entire conversation, heated interactions and public discussions were viewed as norm violations (McLaughlin and Vitak 2012).

Taken together, as prevalence and approval by others influence engagement in the behaviour, it can be proposed that people do not challenge others since they perceive doing so to be unusual and unacceptable on social media.

4) Content-oriented reasons

a) Issue relevance

One reason that individuals choose to remain silent might be the extent to which the content is personally relevant or important to them. Indeed, studies have found that people are more

willing to correct misinformation when the news story is personally relevant to them or their loved ones (Tandoc *et al.* 2020). Another study showed that people skipped past false posts without thoroughly reading them as they did not find them interesting or relevant enough to read fully (Geeng *et al.* 2020).

b) Issue importance

Issue importance also influences people's willingness to speak out publicly on a contentious topic. The greater the perceived importance, the more willing people are to speak out (Moy *et al.* 2001; Gearhart and Zhang 2014). Consequently, it might be argued that people's avoidance of correcting false news can be related to the content's importance, relevance, or appeal.

5) Individual characteristics

Although there are some contextual influences on people's decisions to discuss or confront, individual factors such as demographics (e.g., age, sex, education level) may influence the decision to engage in these conversations. For example, in their study about correction experiences on social media regarding COVID-19, Bode and Vraga (2021a) found that respondents with more education were more likely to engage in correction, and older respondents were less likely to report correcting others.

Personality traits might also influence users' willingness to challenge. The five-factor model of personality describes five dimensions of personality: extraversion, agreeableness, conscientiousness, openness to experience, and neuroticism (McCrae and John 1992). Personality traits influence the frequency and patterns of social interactions in political discussions (Hibbing *et al.* 2011; Gerber *et al.* 2012), commenting on online news (Wu and Atkin 2017) and students' participation in controversial discussions in the classroom (Gronostay 2019).

In the context of politics, there is an association between extraversion and the tendency to discuss politics (Mondak and Halperin 2008; Hibbing *et al.* 2011). As challenging someone requires asking questions or voluntarily providing corrections, extraversion might be positively associated with engaging in conversations to correct misinformation. It may influence people's willingness to approach controversial dialogues on social media. Since agreeable individuals focus on being acceptable in the eyes of others (Graziano and Tobin 2002) and agreeableness is associated with conflict avoidance, it might be negatively related to challenging others. Individuals high in openness to experience are amenable to new ideas and experiences. It can be speculated that openness to experience might be positively associated with approaching conversations to question and learn the perspective of the sharer. Neuroticism describes individuals who are unstable and troubled by negative emotions such as worry and stress (McCrae and John 1992). Therefore, neuroticism might be negatively related to approaching conversations to challenge or correct misinformation.

Perspective-taking and empathic concern could also impact a user's decision to challenge. According to the empathy-altruism hypothesis, empathy for another person generates an altruistic drive to improve that individual's welfare (Batson 1987). As people try to establish a positive self-presentation on SNS (Zhao *et al.* 2008) and sharing misinformation could hurt one's reputation (Altay *et al.* 2019), it can be speculated that users may develop empathy and therefore refrain from challenging misinformation to protect themselves from negative feelings.

6) Technical characteristics

Features and affordances on digital platforms impact users' engagement in several ways. These features can encourage young people to express themselves on political issues (Lane 2020), or affect a user's decision on whether to interact with the content, as they may choose not to engage by using available features rather than commenting (Zhu *et al.* 2017; Wu *et al.* 2020a; Wu *et al.* 2020b). Features like "hide a post", "unfollow", "snooze", and reactions (smiley face, angry face) can be utilised by users as avoidance strategies for not commenting (Wu *et al.* 2020b).

Research has shown that the way in which information is displayed on the interface can have an impact on users' misinformation sharing behaviour (Avram *et al.* 2020; Di Domenico *et al.* 2021). To illustrate, people who are exposed to high engagement metrics (i.e., the numbers of likes and shares) are more likely to like or share false content without verifying it (Avram *et al.* 2020). In addition, when misinformation is presented in a way that the source precedes the message, users are less likely to share it due to a lack of trust (Di Domenico *et al.* 2021). Social media design also affects users' misinformation sharing behaviour (Fazio 2020). Integrating friction into a design, for example, a question to make a user pause and think before sharing information, reduces misinformation sharing (Fazio 2020).

Given that features, affordances, and interface design have an impact on whether to engage with the content or how to engage on SNS, it can also be argued that they may also affect users' decisions to challenge misinformation. The lack of tools provided by the platforms and the way SNS are designed might affect users' tendency to be silent when encountering misinformation.

IV. Conclusion

Despite many studies demonstrating that individuals in cyberspace do not question false information (Chadwick and Vaccari 2019; Tandoc *et al*. 2020; Tully *et al*. 2020; Vicol 2020; Bode and Vraga 2021c) there is much we still do not know about why people remain silent when they encounter misinformation.

By synthesising insights from various bodies of literature, we presented hypotheses about why people refrain from challenging misinformation when they encounter it online. Although there is a commonly held belief that social media is a disinhibited environment where individuals discuss or express anything they like with little concern, studies show that users may feel restrained in some circumstances while expressing their opinions online (Thorson 2014) or correcting misinformation (Tandoc *et al.* 2020). Identifying why people do not engage in conversations to question or correct the content might help devise socio-technical measures to encourage people to challenge misinformation and contribute to mitigating its spread.

Scholars have proposed tools, design considerations, or systems to cultivate constructive discussions in online environments from different fields, e.g., web-based learning environments (Lazonder *et al.* 2003; Yiong-Hwee and Churchill 2007; Hew and Cheung 2008) and political deliberation (Semaan *et al.* 2015; Lane 2020). To the best of our knowledge, no application has been identified for facilitating challenging misinformation online. Given that incorporating design approaches into digital behaviour change interventions is successful in many diverse areas (Elaheebocus *et al.* 2018), design considerations can be extended to encourage users to disagree, question, or correct.

Internet Research

As the persuasive system design model has been employed in a variety of methods to encourage behaviour change (Torning and Oinas-Kukkonen 2009) it offers an opportunity to motivate users to challenge misinformation. Table 1 fleshes out a few design suggestions to illustrate the potential of using PSD strategies. These strategies aim to influence attitudes and behaviour.

Strategy	Definition	Example Implementation
Reduction	The design strategy of reducing complex behaviour into simple tasks	Stickers in instant messaging and SNS are used for many reasons, such as facilitating self-expression, filling the conversation or managing self-impression (Tang *et al*. 2021). Prefabricated stickers with questions to challenge, such as "Did you fact-check this information?" may help users question the content in a quick and impersonal way.
	The design strategy that offers fitting suggestions	Sentence openers are one of the successful techniques used in online learning environments to cultivate students' participation (Lazonder *et al*. 2003; Albertson 2020). A sentence opener is a pre-defined mechanism to start a sentence (Yiong-Hwee and Churchill 2007). "This information is false because..." is an example. A user chooses a sentence opener and adds their comment to finish the argument, such as "the company made a statement that this is false." They may help users in providing well-constructed arguments and challenging misinformation more quickly.
Self-monitoring	The design strategy that enables you to monitor your status or progress.	Affect labelling or labelling one's feelings, helps users to regulate their emotions (Torre and Lieberman 2018). The idea behind using such an approach is to provide insight into users' expressions of emotion. A tone detector is a tool that provides feedback to users about how their comment is likely to sound to someone reading it. As the user writes a comment, the indicator on the scale of emotions begins to form as word choices, style, and punctuation are identified. It may help users to monitor how their comments are likely to sound to someone reading them. This may increase the willingness to challenge misinformation as one of the reasons for refraining from doing so is the fear of seeming aggressive.

Internet Research

Recognition	The design strategy that provides public recognition for performing	A badge can provide public recognition. Users who occasionally correct misinformation can display the badge on their profile. As a result, other users are able to see that this user with the badge has taken the initiative to challenge misinformation on social media. Badges of the type 'Trusted Fact Checker' may motivate users to challenge misinformation more frequently.
Normative influence	The design strategy that displays norms regarding how most people behave and what behaviour they approve	A message that gives information about other users' acceptance and positive attitudes towards correcting misinformation on social media may motivate users. An example is, "Do you know, on this website, 80% of users correct others when they spot misinformation?"
Praise	The design strategy that uses praise as feedback for people's behaviour	A notification or message after correcting misinformation may motivate users. For example, "Your message is the third to dispute this content. Your contribution helps the fight against misinformation."
Rewards	The design strategy that rewards people for performing the target behaviour	A reward such as points after each correction may motivate users to correct misinformation more often. Such points, when accumulated, can translate to a free subscription to a media outlet, e.g., affiliated with where the discussion forums are hosted.

Table I. Design suggestions based on Persuasive System Design (PSD) strategies

1) Research limitations and future research directions

Research limitations

This paper has several limitations. These include a lack of previous research primarily focused on this topic. Therefore, it was not possible to conduct a systematic literature review. Instead, this paper seeks to synthesise relevant literature on related topics and provide potential reasons for users' silence towards misinformation online. It serves as a basis for further studies. We synthesised published literature from a range of fields, including organisational behaviour, communication, human-computer interaction (HCI), psychology, and education. Despite this broad range of topics, we may have missed literature from other research areas. Further, as we searched only published literature, we missed research from the grey literature.

Future research directions

Our identified factors stem from existing literature in different domains. Hence, future research should study whether these reasons are the same as those that prevent users from challenging misinformation in reality. We suggest that future studies expand upon this initial foundation by refining them into sub-factors, investigating other factors that influence the decision not to challenge, and studying whether there are dependencies amongst them. In addition, social

media design can arguably introduce biases (Cemiloglu *et al.* 2021), encourage conformity and hinder individual voices that can challenge what is perceived to be standard behaviour (Zhu *et al*. 2017; Wu *et al*. 2020b; Cemiloglu *et al*. 2021). Future research should investigate such reasons through primary studies. Additionally, research has revealed that open discussion and direct confrontation are more acceptable in Western societies (Morris *et al*. 1998; Friedman *et al*. 2006). A systematic study conducted across cultures on large samples to determine which reasons are common characteristics of human socialisation and which, if any, are distinctive to individual and national experiences may prove useful in the context of challenging misinformation. Another area for future work would involve examining whether challenging behaviour varies according to the individuals' psychometrics and perception of the group and others. For example, research has shown that those who perceive themselves as having less power than the offender are less likely to confront them despite finding their comments inappropriate than those who feel they have higher power (Ashburn-Nardo *et al*. 2014). According to this, the likelihood of challenge may differ depending on the perceived power level of the person they confront (e.g., their boss at work).

Technical solutions have been mainly based on A.I. One direction of research could relate to improving social media design to empower people to speak up when they see misinformation. Our design ideas presented in this paper are only a starting point and intended as hypotheses about how to make people motivated to challenge misinformation through the design of social media. Future research may focus on the actual design and implementation of these ideas. Given the strong link to user experience, methods such as co-design (Sanders 2002; Sanders and Stappers 2014) can be more effective in maintaining the balance between correction requirements and other requirements, including connectedness to others and ease of use. We proposed interventions based on the PSD model. More research is needed to determine whether persuasive techniques are likely to break through the hesitancy and increase the perception of utility regarding challenging misinformation. More research is also required to identify the users' groups concerning their different reactions and preferences to such persuasive interventions. Our identified barriers could also be used to interpret other types of passive online behaviour. Research showed that many individuals who observe instances of racism or prejudice do not attempt to confront the perpetrators (Dickter and Newton 2013). Future research could investigate our hypothesised reasons for online silence beyond the domain of challenging misinformation.

Internet Research

REFERENCES

Acerbi, A., Altay, S. and Mercier, H., 2022. Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School (HKS) Misinformation Review*, 3 (1) pp. 1-15.

Adib, A., Norman, A. and Orji, R., 2021. Persuasive Application for Discouraging Unhealthy Gaming Behaviour, *2021 IEEE 9th International Conference on Serious Games and Applications for Health(SeGAH)* (pp. 1-8): IEEE.

Albertson, B. P., 2020. Promoting Japanese University Students' Participation in English Classroom Discussions: Towards a Culturally-Informed Bottom-Up Approach. *Journal of Pan-Pacific Association of Applied Linguistics*, 24 (1), 45-66.

Alrobai, A., McAlaney, J., Dogan, H., Phalp, K. and Ali, R., 2016. Exploring the requirements and design of persuasive intervention technology to combat digital addiction. *Human-Centered and Error-Resilient Systems Development.* Springer, 130-150.

Altay, S., Hacquin, A.-S. and Mercier, H., 2019. Why do so few people share fake news? It hurts their reputation. *New Media & Society*, 1461444820969893.

Amichai-Hamburger, Y., Gazit, T., Bar-Ilan, J., Perez, O., Aharony, N., Bronstein, J. and Dyne, T. S., 2016. Psychological factors behind the lack of participation in online discussions. *Computers in Human Behavior* 55, 268-277.

Antheunis, M. L., Vanden Abeele, M. M. P. and Kanters, S., 2015. The impact of Facebook use on micro-level social capital: A synthesis. *Societies*, 5 (2), 399-419.

Arif, A., Robinson, J. J., Stanek, S. A., Fichet, E. S., Townsend, P., Worku, Z. and Starbird, K., 2017. A closer look at the self-correcting crowd: Examining corrections in online rumors. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland, Oregon, USA. Association for Computing Machinery. 155–168.

Asch, S. E., 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70 (9), 1.

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A. and Goodwin, S. A., 2014. Do you say something when it's your boss? The role of perpetrator power in prejudice confrontation. *Journal of Social Issues*, 70 (4), 615-636.

Avram, M., Micallef, N., Patil, S. and Menczer, F., 2020. Exposure to social engagement metrics increases vulnerability to misinformation. *Harvard Kennedy School (HKS) Misinformation Review*.

Bandura, A., 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84 (2), 191.

Barlett, C. P., 2015. Anonymously hurting others online: The effect of anonymity on cyberbullying frequency. *Psychology of Popular Media Culture*, 4 (2), 70.

Batson, C. D., 1987. Prosocial Motivation: Is it ever Truly Altruistic? *In:* Berkowitz, L., ed. *Advances in Experimental Social Psychology.* Academic Press, 65-122.

Baumeister, R. F. and Leary, M. R., 1995. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117 (3), 497.

Bautista, J. R., Zhang, Y. and Gwizdka, J., 2021. US Physicians' and Nurses' Motivations, Barriers, and Recommendations for Correcting Health Misinformation on Social Media: Qualitative Interview Study. *JMIR public health and surveillance*, 7 (9), e27715.

Berkowitz, A. D., 2003. Applications of social norms theory to other health and social justice issues. *In:* Perkins, H. W., ed. *The social norms approach to preventing school and college age substance abuse: A handbook for educators, counselors, and clinicians.* Hoboken, NJ: Jossey-Bass/Wiley, 259-279.

Bienefeld, N. and Grote, G., 2012. Silence that may kill: When aircrew members don't speak up and why. *Aviation Psychology and Applied Human Factors*, 2(1), 1–10.

Bode, L. and Vraga, E. K., 2018. See Something, Say Something: Correction of Global Health Misinformation on Social Media. *Health Communication*, 33:9, 1131-1140.

Bode, L. and Vraga, E. K., 2021a. Correction Experiences on Social Media During COVID-19. *Social Media + Society*, 7.

Bode, L. and Vraga, E. K., 2021b. People-powered correction: Fixing misinformation on social media. *The Routledge Companion to Media Disinformation and Populism.* Routledge, 498-506.

Bode, L. and Vraga, E. K., 2021c. Value for Correction: Documenting Perceptions about Peer Correction of Misinformation on Social Media in the Context of COVID-19. *Journal of Quantitative Description: Digital Media*, 1.

Borges-Tiago, T., Tiago, F., Silva, O., Guaita Martinez, J. M. and Botella-Carrubi, D., 2020. Online users' attitudes toward fake news: Implications for brand management. *Psychology & Marketing*, 37 (9), 1171-1184.

Cantarella, M., Fraccaroli, N. and Volpe, R., 2023. Does fake news affect voting behaviour? *Research Policy*, 52 (1).

Cao, X., Khan, A. N., Ali, A. and Khan, N. A., 2020. Consequences of Cyberbullying and Social Overload while Using SNSs: A Study of Users' Discontinuous Usage Behavior in SNSs. *Information Systems Frontiers*, 22 (6), 1343-1356.

Cemiloglu, D., Naiseh, M., Catania, M., Oinas-Kukkonen, H. and Ali, R., 2021. The fine line between persuasion and digital addiction, *Persuasive 2021- International Conference on Persuasive Technology* (pp. 289-307): Springer.

Chadwick, A. and Vaccari, C., 2019. *News sharing on UK social media: Misinformation, disinformation, and correction*. Loughborough: Online Civic Culture Centre, Loughborough University.

Chadwick, A., Vaccari, C. and Hall, N.-A., 2022. *Covid vaccines and online personal messaging: the challenge of challenging everyday misinformation*. Loughborough University.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W. and Starnini, M., 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118 (9).

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F. and Scala, A., 2020. The COVID-19 social media infodemic. *Scientific reports*, 10 (1), 1-10.

Cohen, E. L., Atwell Seate, A., Kromka, S. M., Sutherland, A., Thomas, M., Skerda, K. and Nicholson, A., 2020. To correct or not to correct? Social identity threats increase willingness to denounce fake news through presumed media influence and hostile media perceptions. *Communication Research Reports*, 37 (5), 263-275.

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

Covert, J. M. and Stefanone, M. A., 2018. Does Rejection Still Hurt? Examining the Effects of Network Attention and Exposure to Online Social Exclusion. *Social Science Computer Review*, 38 (2), 170-186.

Darley, J. M. and Latané, B., 1968. Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8 (4p1), 377.

de Oliveira, N. R., Pisa, P. S., Lopez, M. A., de Medeiros, D. S. V. and Mattos, D. M. F., 2021. Identifying fake news on social networks based on natural language processing: trends and challenges. *Information*, 12 (1), 38.

Di Domenico, G., Nunan, D., Sit, J. and Pitardi, V., 2021. Free but fake speech: When giving primacy to the source decreases misinformation sharing on social media. *Psychology & Marketing*, 38 (10), 1700-1711.

Dickter, C. L. and Newton, V. A., 2013. To confront or not to confront: Non-targets' evaluations of and responses to racist comments. *Journal of Applied Social Psychology*, 43, E262-E275.

Djordjevic, M., 2020. Corporate Attempts to Combat Fake News. *Fake News in an Era of Social Media: Tracking Viral Contagion*, 103.

For Peer Review

Dunne, Á., Lawlor, M. A. and Rowley, J., 2010. Young people's use of online social networking sites–a uses and gratifications perspective. *Journal of Research in Interactive Marketing*.

Dyne, L. V., Ang, S. and Botero, I. C., 2003. Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of Management Studies*, 40 (6), 1359-1392.

Edelson, L., Nguyen, M.-K., Goldstein, I., Goga, O., McCoy, D. and Lauinger, T., 2021. Understanding engagement with US (mis) information news sources on Facebook, *IMC '21: Proceedings of the 21st ACM Internet Measurement Conference* (pp. 444-463).

Elaheebocus, S. M. R. A., Weal, M., Morrison, L. and Yardley, L., 2018. Peer-based social media features in behavior change interventions: systematic review. *Journal of medical Internet research*, 20 (2), e8342.

Fassinger, P. A., 1995. Understanding classroom interaction: Students' and professors' contributions to students' silence. *The Journal of Higher Education*, 66 (1), 82-96.

Fazio, L., 2020. Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, 1 (2).

Fingerman, K. L., 2009. Consequential strangers and peripheral ties: The importance of unimportant relationships. *Journal of Family Theory & Review*, 1 (2), 69-86.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M. and Kainbacher, M., 2011. The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137 (4), 517.

Friedman, R., Chi, S.-C. and Liu, L. A., 2006. An expectancy model of Chinese–American differences in conflict-avoiding. *Journal of International Business Studies*, 37 (1), 76-91.

Gallrein, A. M. B., Bollich-Ziegler, K. L. and Leising, D., 2019. Interpersonal feedback in everyday life: Empirical studies in Germany and the United States. *European Journal of Social Psychology*, 49 (1), 1-18.

Galvão, J., 2021. COVID-19: the deadly threat of misinformation. *The Lancet Infectious Diseases*, 21 (5), e114.

Gearhart, S. and Zhang, W., 2014. Gay bullying and online opinion expression. *Social Science Computer Review*, 32 (1), 18-36.

Geeng, C., Yee, S. and Roesner, F., 2020. Fake News on Facebook and Twitter: Investigating how people (don't) investigate. *Conference on Human Factors in Computing Systems Proceedings*, 1-14.

Gerber, A. S., Huber, G. A., Doherty, D. and Dowling, C. M., 2012. Disagreement and the avoidance of political discussion: Aggregate relationships and differences across personality traits. *American Journal of Political Science*, 56 (4), 849-874.

Gimpel, H., Heger, S., Olenberger, C. and Utz, L., 2021. The effectiveness of social norms in fighting fake news on social media. *Journal of Management Information Systems*, 38 (1), 196-221.

Graziano, W. G. and Tobin, R. M., 2002. Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality*, 70 (5), 695-728.

Gronostay, D., 2019. To argue or not to argue? The role of personality traits, argumentativeness, epistemological beliefs and assigned positions for students' participation in controversial political classroom discussions. *Unterrichtswissenschaft*, 47 (1), 117-135.

Hampton, K. N., Rainie, H., Lu, W., Dwyer, M., Shin, I. and Purcell, K., 2014. *Social media and the'spiral of silence'*. Pew Research Center.

Hew, K. F. and Cheung, W. S., 2008. Attracting student participation in asynchronous online discussions: A case study of peer facilitation. *Computers & Education*, 51 (3), 1111-1124.

Hibbing, M. V., Ritchie, M. and Anderson, M. R., 2011. Personality and political discussion. *Political Behaviour*, 33 (4), 601-624.

Ho, S. S. and McLeod, D. M., 2008. Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35 (2), 190-207.

Hudson, J. M. and Bruckman, A. S., 2004. The Bystander Effect: A Lens for Understanding Patterns of Participation. *Journal of the Learning Sciences*, 13 (2), 165-195.

Jaworski, A. and Sachdev, I., 1998. Beliefs about silence in the classroom. *Language Education*, 12 (4), 273-292.

Joinson, A. N., 2008. Looking at, looking up or keeping up with people? Motives and use of Facebook, *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1027-1036).

Katie Elson, A., 2018. Getting acquainted with social networks and apps: combating fake news on social media. *Library Hi Tech News*, 35 (3), 1-6.

Kawachi, I. and Berkman, L. F., 2001. Social ties and mental health. *Journal of Urban health*, 78 (3), 458-467.

Kim, A., Moravec, P. L. and Dennis, A. R., 2019. Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems*, 36 (3), 931-968.

Kim, J., Tabibian, B., Oh, A., Schölkopf, B. and Gomez-Rodriguez, M., 2018. Leveraging the crowd to detect and reduce the spread of fake news and misinformation, *WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 324-332).

Koo, A. Z.-X., Su, M.-H., Lee, S., Ahn, S.-Y. and Rojas, H., 2021. What Motivates People to Correct Misinformation? Examining the Effects of Third-person Perceptions and Perceived Norms. *Journal of Broadcasting & Electronic Media*, 65 (1), 111-134.

Koutamanis, M., Vossen, H. G. M. and Valkenburg, P. M., 2015. Adolescents' comments in social media: Why do adolescents receive negative feedback and who is most at risk? *Computers in Human Behavior*, 53, 486-494.

Lane, D. S., 2020. Social media design for youth political expression: Testing the roles of identifiability and geo-boundedness. *New Media & Society*, 22 (8), 1394-1413.

Lapinski, M. K. and Rimal, R. N., 2005. An explication of social norms. *Communication theory*, 15 (2), 127-147.

Lasorsa, D. L. J. J. Q., 1991. Political outspokenness: Factors working against the spiral of silence. *Journalism Quarterly*, 68 (1-2), 131-140.

Lazonder, A. W., Wilhelm, P. and Ootes, S. A. W., 2003. Using sentence openers to foster student interaction in computer-mediated learning environments. *Computers & Education*, 41 (3), 291-308.

Leary, M. R. and Kowalski, R. M., 1990. Impression management: A literature review and two-component model. *Psychological Bulletin*, 107 (1), 34.

Luarn, P. and Hsieh, A.-Y., 2014. Speech or silence. *Online Information Review*, 38 (7), 881-895.

Lutzke, L., Drummond, C., Slovic, P. and Árvai, J., 2019. Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change Part A: Human & Policy Dimensions*, 58, N.PAG-N.PAG.

Lyons, B., 2017. Insidiously Trivial: How Political Memes Drive Down Corrective Intent. https://www.academia.edu/33740160/Insidiously_Trivial_How_Political_Memes_Drive_Down_Corrective_Intent: University of Exeter.

Margolin, D. B., Hannak, A. and Weber, I., 2018. Political Fact-Checking on Twitter: When Do Corrections Have an Effect? *Political Communication*, 35:2, 196-219.

Marwick, A. E. and Boyd, D., 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13 (1), 114-133.

McClain, C., 2021. *American Trend Panel*.

McCrae, R. R. and John, O. P., 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60 (2), 175-215.

McLaughlin, C. and Vitak, J., 2012. Norm evolution and violation on Facebook. *New Media & Society*, 14 (2), 299-315.

Internet Research

Milliken, F. J., Morrison, E. W. and Hewlin, P. F., 2003. An exploratory study of employee silence: Issues that employees don't communicate upward and why. *Journal of Management Studies*, 40 (6), 1453-1476.

Mishra, A. and Samu, S., 2021. Impact of fake news on social image perceptions and consumers' behavioral intentions. *Journal of Consumer Marketing*.

Misra, G., Arachchilage, N. A. G. and Berkovsky, S., 2017. Phish phinder: a game design approach to enhance user confidence in mitigating phishing attacks. *International Symposium on Human Aspects of Information Security & Assurance (HAISA 2017)*.

Mondak, J. J. and Halperin, K. D., 2008. A framework for the study of personality and political behaviour. *British Journal of Political Science*, 38 (2), 335-362.

Morris, M. W., Williams, K. Y., Leung, K., Larrick, R., Mendoza, M. T., Bhatnagar, D., Li, J., Kondo, M., Luo, J.-L. and Hu, J.-C., 1998. Conflict management style: Accounting for cross-national differences. *Journal of international business studies*, 29 (4), 729-747.

Morrison, E. W. and Milliken, F. J., 2000. Organizational silence: A barrier to change and development in a pluralistic world. *The Academy of Management Review*, 25 (4), 706-725.

Moy, P., Domke, D. and Stamm, K., 2001. The Spiral of Silence and Public Opinion on Affirmative Action. *Journalism & Mass Communication Quarterly*, 78 (1), 7-25.

Nakov, P., 2020. Can We Spot the" Fake News" Before It Was Even Written? *arXiv preprint arXiv:2008.04374*.

Noelle-Neumann, E., 1974. The Spiral of Silence A Theory of Public Opinion. *Journal of Communication*, 24 (2), 43-51.

Nonnecke, B. and Preece, J., 2001. Why Lurkers Lurk, *America's Conference on Information Systems (AMCIS)*.

Oinas-Kukkonen, H. and Harjumaa, M., 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24 (1), 28.

Orji, R. and Moffatt, K., 2018. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Informatics Journal*, 24 (1), 66-91.

Oyibo, K. and Vassileva, J., 2019. Investigation of persuasive system design predictors of competitive behavior in fitness application: A mixed-method approach. *Digital Health*, 5, 2055207619878601.

Ozturk, P., Li, H. and Sakamoto, Y., 2015. Combating rumor spread on social media: The effectiveness of refutation and warning, *2015 48th Hawaii international conference on system sciences* (pp. 2406-2414): IEEE.

Paliszkiewicz, J. and Mądra-Sawicka, M., 2016. Impression Management in Social Media: The Example of LinkedIn. *Management*, 11 (3).

Papacharissi, Z., 2002. The virtual sphere: the internet as a public sphere. *New media & Society*, 4 (1), 9-27.

Park, C. Y., Mendelsohn, J., Field, A. and Tsvetkov, Y., 2022. VoynaSlov: A Data Set of Russian Social Media Activity during the 2022 Ukraine-Russia War. *arXiv preprint arXiv:2205.12382*.

Patchin, J. W. and Hinduja, S., 2006. Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, 4 (2), 148-169.

Pinder, C. C. and Harlos, K. P., 2001. Employee silence: Quiescence and acquiescence as responses to perceived injustice. *Research in personnel and human resources management.* Emerald Group Publishing Limited.

Pourghomi, P., Safieddine, F., Masri, W. and Dordevic, M., 2017. How to stop spread of misinformation on social media: Facebook plans vs. right-click authenticate approach, *2017 International Conference on Engineering & MIS (ICEMIS)* (pp. 1-8): IEEE.

Powers, E., Koliska, M. and Guha, P., 2019. "Shouting Matches and Echo Chambers": Perceived Identity Threats and Political Self-Censorship on Social Media. *International Journal of Communication*, 13, 20.

Internet Research

Rimal, R. N. and Real, K., 2003. Understanding the influence of perceived norms on behaviors. *Communication Theory*, 13 (2), 184-203.

Rocca, K. A., 2010. Student participation in the college classroom: An extended multidisciplinary literature review. *Communication Education*, 59 (2), 185-213.

Rohman, A., 2021. Counteracting Misinformation in Quotidian Settings, *International Conference on Information* (pp. 141-155): Springer.

Sanders, E. B.-N., 2002. From user-centered to participatory design approaches. *Design and the social sciences.* CRC Press, 18-25.

Sanders, E. B. N. and Stappers, P. J., 2014. Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign*, 10 (1), 5-14.

Semaan, B., Faucett, H., Robertson, S. P., Maruyama, M. and Douglas, S., 2015. Designing political deliberation environments to support interactions in the public sphere, *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3167-3176).

Shearer, E. and Mitchell, A., 2021. *News Use Across Social Media Platforms 2017*.

Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J. and Cranor, L. F., 2013. The post that wasn't: exploring self-censorship on facebook. *Proceedings of the 2013 conference on Computer supported cooperative work*, San Antonio, Texas, USA. Association for Computing Machinery. 793–802. Available from: https://doi.org/10.1145/2441776.2441865 [Accessed

Slonje, R., Smith, P. K. and Frisén, A., 2013. The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior*, 29 (1), 26-32.

Steen-Johnsen, K. and Enjolras, B., 2016. The Fear of Offending: Social Norms and Freedom of Expression. *Society*, 53 (4), 352-362.

Stroud, N. J., Van Duyn, E. and Peacock, C., 2016. News commenters and news comment readers. *Engaging News Project*, 1-21.

Suler, J., 2004. The Online Disinhibition Effect. *CyberPsychology & Behavior*, (3), 321-326.

Sun, Y., Oktavianus, J., Wang, S. and Lu, F., 2021. The role of influence of presumed influence and anticipated guilt in evoking social correction of COVID-19 misinformation. *Health Communication*, 1-10.

Tanaka, Y., Sakamoto, Y. and Matsuka, T., 2013. Toward a Social-Technological System that Inactivates False Rumors through the Critical Thinking of Crowds [Conference]. *2013 46th Hawaii International Conference on System Sciences*. IEEE. 649-658. Available from: https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.6479912&site=eds-live&scope=site [Accessed

Tandoc, E. C., Lim, D. and Ling, R., 2020. Diffusion of disinformation: How social media users respond to fake news and why. *Journalism*, 21 (3), 381-398.

Tang, Y., Hew, K. F., Herring, S. C. and Chen, Q., 2021. (Mis) communication through stickers in online group discussions: A multiple-case study. *Discourse & Communication*, 15 (5), 582-606.

Thorson, K., 2014. Facing an uncertain reception: Young citizens and political interaction on Facebook. *Information, Communication & Society*, 17 (2), 203-216.

Torning, K. and Oinas-Kukkonen, H., 2009. Persuasive system design: state of the art and future directions, *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology* (pp. 1-8).

Torre, J. B. and Lieberman, M. D., 2018. Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review*, 10 (2), 116-124.

Tully, M., Bode, L. and Vraga, E. K., 2020. Mobilizing Users: Does Exposure to Misinformation and Its Correction Affect Users' Responses to a Health Misinformation Post? *Social Media + Society*, 6.

Urbaniak, R., Ptaszyński, M., Tempska, P., Leliwa, G., Brochocki, M. and Wroczyński, M., 2022. Personal attacks decrease user activity in social networking platforms. *Computers in Human Behavior*, 126, 106972.

Vicol, D. O., 2020. *Who is most likely to believe and to share misinformation?* https://fullfact.org/media/uploads/who-believes-shares-misinformation.pdf.
Visentin, M., Pizzi, G. and Pichierri, M., 2019. Fake news, real problems for brands: The impact of content truthfulness and source credibility on consumers' behavioral intentions toward the advertised brands. *Journal of Interactive Marketing*, 45 (1), 99-112.
Visentin, M., Tuan, A. and Di Domenico, G., 2021. Words matter: How privacy concerns and conspiracy theories spread on twitter. *Psychology & Marketing*, 38 (10), 1828-1846.
Vosoughi, S., Roy, D. and Aral, S., 2018. The spread of true and false news online. *Science*, 359 (6380), 1146-1151.
Vraga, E. K. and Bode, L., 2017. Using expert sources to correct health misinformation in social media. *Science Communication*, 39 (5), 621-645.
Vraga, E. K. and Bode, L., 2018. I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21 (10), 1337-1353.
Vraga, E. K., Thorson, K., Kligler-Vilenchik, N. and Gee, E., 2015. How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior*, 45, 281-289.
Walter, N., Brooks, J. J., Saucier, C. J. and Suresh, S., 2021. Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication*, 36 (13), 1776-1784.
Walter, N. and Murphy, S. T., 2018. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85:3, 423-441.
Walther, J. B., 1992. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19 (1), 52-90.
Weinstein, E. C., 2014. The personal is political on social media: Online civic expression patterns and pathways among civically engaged youth. *International journal of communication*, 8, 24.
Westerwick, A., Johnson, B. K. and Knobloch-Westerwick, S., 2017. Confirmation biases in selective exposure to political online information: Source bias vs. content bias. *Communication Monographs*, 84 (3), 343-364.
Widyasari, Y. D. L., Nugroho, L. E. and Permanasari, A. E., 2019. Persuasive technology for enhanced learning behavior in higher education. *International Journal of Educational Technology in Higher Education*, 16 (1), 1-16.
Williams, K. D., Cheung, C. K. and Choi, W., 2000. Cyberostracism: effects of being ignored over the Internet. *Journal of Personality*

Social Psychology, 79 (5), 748.
Williams, K. D. and Nida, S. A., 2011. Ostracism: Consequences and coping. *Current Directions in Psychological Science*, 20 (2), 71-75.
Williams, K. D. and Sommer, K. L., 1997. Social ostracism by coworkers: Does rejection lead to loafing or compensation? *Personality and Social Psychology Bulletin*, 23 (7), 693-706.
Wu, T.-Y. and Atkin, D., 2017. Online News Discussions: Exploring the Role of User Personality and Motivations for Posting Comments on News. *Journalism & Mass Communication Quarterly*, 94 (1), 61-80.
Wu, T.-Y., Oeldorf-Hirsch, A. and Atkin, D., 2020a. A click is worth a thousand words: Probing the predictors of using click speech for online opinion expression. *International Journal of Communication*, 14, 20.
Wu, T.-Y., Xu, X. and Atkin, D., 2020b. The alternatives to being silent: exploring opinion expression avoidance strategies for discussing politics on Facebook. *Internet Research*.
Xiao, X., 2022. Let's verify and rectify! Examining the nuanced influence of risk appraisal and norms in combatting misinformation. *New Media & Society*, 14614448221104948.
Yiong-Hwee, T. and Churchill, D., 2007. Using Sentence Openers to Support Students' Argumentation in an Online Learning Environment. *Educational Media International*, 44 (3), 207-218.

You, L. and Lee, Y.-H., 2019. The bystander effect in cyberbullying on social network sites: Anonymity, group size, and intervention intentions. *Telematics and Informatics*, 45, 101284.

Zhao, H., Fu, S. and Chen, X., 2020. Promoting users' intention to share online health articles on social media: The role of confirmation bias. *Information Processing & Management*, 57 (6), 102354.

Zhao, S., Grasmuck, S. and Martin, J., 2008. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 24 (5), 1816-1836.

Zhu, Q., Skoric, M. and Shen, F., 2017. I shield myself from thee: Selective avoidance on social media during political protests. *Political Communication*, 34 (1), 112-131.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S. and Quattrociocchi, W., 2017. Debunking in a World of Tribes. *PLoS ONE*, 12.

Internet Research