Can we re-design social media to persuade people to challenge misinformation? An exploratory study

Selin Gurgun[1], Emily Arden-Close[1], John McAlaney[1], Keith Phalp[1] and Raian Ali[2]

[1] Faculty of Science and Technology, Bournemouth University, Poole, UK
{sgurgun, eardenclose, jmcalaney, kphalp} @ bournemouth.ac.uk
[2] College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
raali2@hbku.edu.qa

Abstract. Persuasive design techniques have often been presented where the desired behaviour is primarily within personal boundaries, e.g., one's own health and learning. Limited research has been conducted on behaviours that require exposure to others, including correcting, confronting mistakes and wrongdoing. Challenging misinformation in others' posts online is an example of such social behaviour. This study draws on the main persuasive system design models and principles to create interfaces on social media to motivate users to challenge misinformation. We conducted a questionnaire (with 250 participants from the UK) to test the influence of these interfaces on willingness to challenge and how age, gender, personality traits, perspective-taking and empathy affected their perception of the persuasiveness of the interfaces. Our proposed interfaces exemplify seven persuasive strategies: reduction, suggestion, self-monitoring, recognition, normative influence, tunneling and liking. Most participants thought existing social media did not provide enough techniques and tools to challenge misinformation. While predefined question stickers (suggestion), private commenting (reduction), and thinking face reactions (liking) were seen as effective ways to motivate users to challenge misinformation, sentence openers (tunneling) was seen as the least influential. Increasing age and perspective taking were associated with increased likelihood of perceived persuasiveness and increasing openness to experience was associated with a reduction in the likelihood of perceived persuasiveness for "predefined question stickers". Increasing openness to experience was associated with increased likelihood of perceived persuasiveness for "thinking face reaction", while increasing age was associated with a reduction in the likelihood of perceived persuasiveness for "private commenting".

Keywords: Persuasive system, misinformation, fake news, online social behaviour, social media design,

1   Introduction

The growing proliferation of misinformation has raised interest in devising ways to combat it. Since social media plays a paramount role in disseminating misinformation [1, 2], technology-based solutions to detect and mitigate misinformation are being proposed and used. Such solutions rely on artificial intelligence (AI) and machine learning

[3, 4]. Despite offering promise, the existence of misinformation in recent events such as the COVID-19 pandemic [5] and Ukraine-Russia War [6] demonstrates that these techniques have limitations. In addition to algorithmic approaches provided by digital platforms and legislative measures taken by governments and social media platforms, individual attempts to correct misinformation have also been effective [7].

Although challenging misinformation is effective in reducing the spread of misinformation [7, 8], research shows that when people encounter misinformation in the online sphere, they tend to remain silent [9-12]. However, challenging others helps the poster correct or delete the shared information and helps others in the network adjust their perceptions about the content [13]. Therefore, users' silence contributes, albeit unintentionally, to the propagation of misinformation.

Studies on why people remain silent when they encounter misinformation indicate that users have interpersonal, intrapersonal and institutional concerns such as fear of being harassed, fear of conflict or lack of institutional support [10, 14]. Gurgun et al. [15] identified six reasons that might impact users' decisions to challenge others who post misinformation: self-oriented, relationship-oriented, others-oriented, content-oriented, individual characteristics, and technical factors. While it is necessary to overcome the barriers that hinder users from challenging misinformation, it is also worthwhile to design social media platforms that are more persuasive and encourage users to challenge misinformation.

In a digital environment, information systems may be designed to influence behaviour change [16]. Such systems are often referred to as persuasive systems [17] or persuasive technology [18]. These systems address behaviour change in areas where the intended behaviour is primarily within personal boundaries such as one's own wellbeing [19] and academic performance [20]. However, to the best of our knowledge, a few studies have addressed behaviours including interactions with other people. One study found that persuasive technology can promote acceptable workplace behaviour and etiquette [21]. As incorporating persuasive techniques into social media has been shown to lead to greater behaviour change and more engagement [22, 23] they could be applied to motivate users to challenge misinformation.

In this paper, we aim to develop an understanding of which persuasive techniques may aid users in challenging others in an environment where algorithms and a homogeneous network impede exposure to adverse opinions [24-26]. Given the lack of research on users' reluctance to challenge misinformation on social media, we aimed to provide a starting point and investigate potential solutions based on persuasive design principles, linking them to user characteristics such as personality traits, perspective-taking, and empathy. While we do not claim completeness in our coverage of design techniques and user characteristics, our choices were informed by theory. Personality traits have been shown to influence the willingness to participate in discussions, e.g. extroverts are typically more willing to do so [27]. In addition to that, evidence from the literature suggests that there is a significant relationship between persuasive strategies and personality traits [28, 29] To the best of our knowledge, no attempt has been made to examine the associations between perspective-taking, empathic concern and persuasive techniques. Perspective-taking refers to the ability to adopt the perspective of other people [30]. According to research, perspective-taking influences attributional

thinking and perceptions of others. For example, those who with a higher perspective-taking tendency made the same attributions to the target as they would have had they been in the same situation [31]. As giving negative feedback, disagreeing and confronting are related to the perception of others (e.g., whether they perceive it as futile or harmful to their relationship), we also examine whether perspective-taking has an association with the perception of the persuasiveness of the proposed interfaces. Empathic concern signifies the development of emotions of compassion and concern for those experiencing negative events [30]. Individuals with higher level of empathic concern are more likely to acquire positive attitudes toward being a decent and moral person and not to harm others [32]. Therefore, since empathy might be a useful component for challenging misinformation, we also investigated whether the interfaces we presented are associated with users' empathic concern. This insight may help for the future interface design to motivate users to challenge misinformation. In this paper, we aim to answer the following research questions:

RQ1: To what extent do users think that current social media design helps them to challenge misinformation?

RQ2: Can we increase users' willingness to challenge misinformation on social media through introducing persuasive interfaces?

RQ3: Do age, gender, personality traits, perspective-taking, and empathy have an impact on users' acceptance of these interfaces?

2 Method

Our work aims to investigate whether we can design new persuasive interfaces in social media to increase users' willingness to challenge misinformation. To achieve this goal, first, based on relevant literature, we explored effective persuasive tools that aid in increasing engagement or participation in the online social sphere. To stay within a manageable scope, we selected seven strategies (Table 1). Then, we created a high-fidelity prototype of a possible implementation of each strategy in the context of challenging misinformation. We then conducted a survey to assess users' perceptions of the influence of each prototype on their willingness to challenge misinformation.

2.1 Prototype Design

In order to create interfaces for challenging misinformation, we started with reviewing the pertinent literature regarding the tools and techniques that have a positive impact on motivating people to engage more in online discussion environments. We reviewed literature regarding the design recommendations suggested for promoting healthy discussions [33], strengthening engagement in discussions [34] and increasing participation in online communities [35]. We selected seven techniques based on their practicality and possibility to embed within social media design with no costly requirement. The interfaces have been deemed face valid through a combination of expert review and pilot testing with users. Descriptions of the techniques applied in this study, which were

adapted from [36] are shown in Table 1 and prototypes are available on Open Science Framework (see https://osf.io/3x74c).

Table 1. Persuasive Strategies, corresponding descriptions and implementations

Strategy	Description	Our Implementation
Reduction	A strategy of reducing complex behaviour into simple tasks	**Private Commenting (PC):** This option makes private messaging easier. Instead of copying the link to the post in the private message section to refer to it, users can comment privately on the post directly by clicking the "send privately" button.
Suggestion	A strategy that offers fitting suggestions	**Predefined Question Stickers (PQS):** A sticker set with pre-written questions for users to choose from when they want to challenge, e.g. stickers with labels like "what is your source"
Self-monitoring	A strategy that enables monitoring of one's own status or progress.	**Tone Detector (TD):** An emotional scale mood indicator that enables users to visualise how their comments are likely to sound to someone reading them. As the user composes a comment, the indicator on the scale starts to move based on the word selections, writing style and punctuation, e.g., aggressive vs friendly.
Recognition	A strategy that provides public acknowledgement for following certain behaviour	**"Fact Checker" Badge (FCB):** A badge that provides public recognition for users who correct misinformation occasionally.
Normative influence	A strategy that displays norms regarding how most people behave and what behaviour they approve of.	**Social Norm Message (SNM):** A pop-up prompt regarding other users' acceptance and positive disposition towards correcting misinformation on social media.
Tunneling	A strategy that guides users or provides means to persuade.	**Sentence openers (SO):** Pre-generated sentence openers to guide users to challenge misinformation, e.g. "My argument is…"
Liking	A strategy that highlights the persuasiveness of visually attractive	**Thinking Face Reaction (TFR):** A fun and appealing reaction that implies questioning the content.

Recommended techniques were visualised to mimic Facebook's current user interface in the survey. The content displayed to participants was about an asteroid that will be possibly hitting Earth. This news was a misinformation article that appeared in CNN's iReport news hub in 2014[1] and was widely shared. We presented the interfaces

---

[1] https://www.cnet.com/science/cnn-posts-asteroid-to-hit-earth-article-people-take-it-seriously/

to participants as a scenario. Before showing the interfaces, we informed the respondents about the scenario that the news is false, as this study was not to determine whether individuals can discern misinformation but rather to assess perceptions of the persuasive techniques. Additionally, care was taken not to include any political or social issues as these topics may influence users' decisions to challenge the misinformation regardless of the persuasive technique used [37, 38]. To make the scenario more realistic and prevent participants from perceiving the account as anonymous, we intentionally did not obscure the name and photo of the Facebook user sharing the content. The account holder was named Alex to avoid gender confounding as it is a gender-neutral name. Facebook was selected as the social network site for this study as it is the most used online social network worldwide with approximately 2.91 billion monthly active users [39] across all age groups [40]. The questions were about challenging an acquaintance, as previous research showed that it is more difficult to debate or challenge weak ties than strong ties [41]. Before presenting the interfaces we also explained that Alex is a contact who is known but not as close as a friend or family member, which also controlled for familiarity and personal ties.

Below, we explain each of the interfaces that we used and how they exemplify the persuasive design principles.

Private commenting (Reduction) A private mode of communication such as direct messaging (DM) or private messaging (PM) can be preferred by users who hesitate to counteract misinformation due to fear of embarrassing the sharer [42]. Although Facebook provides private communication channels (e.g., Facebook Messenger), private commenting is proposed as it simplifies the challenging process. Private messaging requires effort. In contrast, private commenting allows users to engage with the content while scrolling through their feeds.

Pre-defined question stickers (Suggestion) By suggesting to users some questions to challenge, we aimed to aid users who struggle with starting challenging misinformation. We developed three exemplar questions based on Toulmin's model of argumentation [43] which we implemented as stickers. As stickers are impersonal responses, they may alleviate the psychological and social costs associated with confronting, questioning and challenging by typing directly [44].

Tone detector (Self-monitoring) Self-monitoring is one of the most popular persuasive strategies [45] and it helps people become more self-aware about their behavioural patterns [46] which also helps them to self-regulate [47]. One potential reason people do not challenge misinformation is that they want to avoid being viewed as aggressive [15]. A tone detector guides users to improve their self-awareness regarding how they sound to others, which would likely be helpful for those who are concerned about how their comments come across.

Fact Checker Badge (Recognition) In addition to providing public recognition, badges can function as incentives for people who challenge misinformation. SNS like Foursquare or StackOverflow have also used badges as a tool to encourage users to increase their level of engagement [48].

Social Norm Messages (Normative Influence) Normative social influence can be successfully used in persuasive technologies to alter existing behaviour [49, 50]. Social norms refer to the beliefs that individuals hold about others in a social group and what others in that social group do or think [51-53]. People are more likely to engage in a behaviour if they perceive it to be commonly accepted by others [54]. Taking that into account we proposed a hint prompt that provides information regarding injunctive norms (what most people approve of).

Sentence openers (Tunneling) Tunneling is defined as guiding a user through a complex experience [36]. We used sentence openers to lead users to complete the sentences. In many cases, it is rhetoric rather than facts that causes contradictory opinions [55]. Sentence starters have been shown to facilitate expression of disagreement [56] and enhance the quality of online discussions [57] as they can steer thoughts and conversation in such a way that they keep conversations grounded [58].

Thinking face reaction (Liking) Facebook launched the "Like" button to enable users to show their affection for the content and introduced the "Reactions" feature to provide users with more ways to express their reactions to content in a fast and easy way [59]. The reactions feature has positively influenced engagement levels on Facebook [60]. As reactions are intended to be a way to express appreciation for the content with one-click, we used thinking face reactions to motivate users to challenge [61].

2.2   Data Collection

The survey was conducted online using Qualtrics, an online survey design platform. Before survey completion, participants were informed about the study objectives and asked to provide their consent to participate. They were provided with information regarding the confidentiality of the data, their freedom to participate and the right to withdraw from the study as well as their access to the study findings. Data was collected between 31st May and 7th July 2022 through Prolific™ (www.prolific.co), an established online participant recruitment platform for research studies.

2.3   Participants

250 adults living in the UK completed the survey and received payment (around £4). The following inclusion criteria were used to recruit potential participants: 1) aged 18 years or older; 2) using Facebook with an authentic identity 3) encountered misinformation on Facebook. Of those respondents who answered the demographic questions, 57.2% (143) were female, 41.6% (104) male, and 1.2% (3) non-binary. 17.8% (44) were aged 18 to 24 years, 37.2% (92) were between 25-34, 26.7% (66) between 35-44, and 18.2% (45) over 45 years. Most respondents (62.8%, 157) had at least a university degree, 22.8% (57) had completed secondary education and 14.4% (36) had not completed secondary education.

2.4 Measures

Demographic Characteristics
Participants answered questions about their age, gender and educational level.

Perceived prevalence of existing tools to challenge misinformation
Participants were asked to rate the extent to which they think social media provides tools or ways for users to challenge misinformation on a seven-point scale (1= None at all 7= A great deal) (M = 3.37, SD = 1.6).

Perceived persuasiveness on willingness to challenge
For this study, we developed seven prototypes informed by PSD and presented the participants eight high-fidelity prototypes including the standard comment box (SCB). Participants were asked how much each interface influenced their willingness to challenge misinformation on a seven-point scale (1= Far too little 7= Far too much). Prototypes were presented in random order for each respondent.

Personality traits
The 10-item Big-Five inventory (BFI-10) [62] was used to assess the traits of extraversion, agreeableness, openness to experience, conscientiousness and neuroticism-stability. Each trait is measured by two items on a five-point Likert scale (1 = Strongly disagree to 5 = Strongly agree). Higher scores imply higher levels of each personality trait. The BFI-10 showed good reliability and validity [62].

Perspective-taking and empathic concern
Perspective taking and empathic concern were assessed with the seven-item perspective-taking and seven-item empathy subscales from the Interpersonal Reactivity Index (IRI) which assesses level of empathy and individuals' capacity to understand and feel the emotions of others [30]. It is widely used to examine the impact of empathy on various outcomes, including prosocial behaviour [63] and intergroup relations [64]. The perspective-taking subscale consisted of seven items ($\alpha = .78$) such as ''I try to look at everybody's side of a disagreement before I make a decision''). Empathy subscale consisted of seven items ($\alpha = .81$) such as "When I see someone being taken advantage of, I feel kind of protective toward them."). The items are answered using a 5-point rating scale ranging from on a 0 (does not describe me well), to 4 (describes me very well). To get a single score for each participant for each construct, we averaged the answers to the items.

2.5 Data Analysis

Data were analysed using SPSS version 28. Descriptive statistics were used to report demographic information and the extent to which users think social media provides them with tools or techniques to counteract misinformation. As the data was not normally distributed, non-parametric tests were used. Spearman's rank-order correlation

was used to analyse the association between continuous and ordinal data. A Wilcoxon signed-rank test was conducted to investigate differences in willingness to challenge misinformation between the status quo (standard comment box) and PSD-based interfaces. Interfaces were considered persuasive when users rated them as more influential than the standard comment box in motivating challenging misinformation. Binomial logistic regression analyses were performed to determine whether age, gender, personality traits, perspective-taking and empathy influence willingness to challenge misinformation.

3 Results

3.1 Social media tools to challenge misinformation

Most participants (58%) agreed that current social media platforms do not provide enough tools or ways to challenge misinformation. Only 26.8% believed that tools social media provides for challenging misinformation are prevalent. A Spearman's rank-order correlation was run to assess the relationship between age and evaluation of the tools on social media to challenge misinformation. Age was negatively correlated with agreement that social media offers tools to challenge r_s (248) = -.153, p < .05. Older participants were less likely to think that social media provide tools for challenging misinformation.

3.2 Comparing the perceived persuasiveness of each interface

A Wilcoxon signed-rank test asssessed the effect of the interfaces presented on users' willingness to challenge misinformation. As shown in Table 2, "predefined question stickers", "thinking face reaction" and "private commenting" elicited a statistically significant positive difference ($z = 2.43$, $p < .05$, $z = 5$, $p < .001$ and $z = 7.2$, $p < .001$ respectively) and "sentence openers" showed a statistically significant negative difference ($z = -2.7$, $p < .05$) in their influence on willingness to challenge misinformation compared to the standard comment box. Excepting sentence openers, most participants rated PSD-informed implementations as having more influence than the standard comment box.

Table 2. The influence on willingness to challenge for each presented interface versus standard comment box (N = 250)

Predefined Stickers (PQS)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	80[a]	110.84	8867.50	-2.430[*]		< .05
Positive Ranks	129[b]	101.38	13077.50			
Ties	41[c]					
Thinking Face Reaction (TFR)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	65[d]	93.04	6047.50	-5.005[*]		< .001
Positive Ranks	136[e]	104.81	14253.50			
Ties	49[f]					
Private Commenting (PC)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	47[g]	99.67	4684.50	-7.239[*]		< .001
Positive Ranks	162[h]	106.55	17260.50			
Ties	41[i]					
Sentence Openers (SO)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	102[j]	96.22	9814.00	-2.714[¶]		< .05
Positive Ranks	76[k]	80.49	6117.00			
Ties	72[l]					
Fact Checker Badge (FCB)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	93[m]	102.62	9543.50	-.139[*]		0.89
Positive Ranks	103[n]	94.78	9762.50			
Ties	54[o]					
Social Norm Message (SNM)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	76[p]	111.68	8488.00	-1.821[*]		0.06
Positive Ranks	123[q]	92.78	11412.00			
Ties	51[r]					
Tone Detector (TD)	N	Mean Ranks	Sum of Ranks	Z	P	
Negative Ranks	91[s]	99.47	9051.50	-1.344[*]		0.17
Positive Ranks	110[t]	102.27	11249.50			
Ties	49[u]					

a. PQS < SCB   b. PQS > SCB   c. PQS = SCB   d. TFR < SCB   e. TFR > SCB   f. TFR = SCB

g. PC < SCB   h. PC > SCB   i. PC = SCB   j. SO < SCB   k. SO > SCB   l. SO = SCB

m. FCB < SCB   n. FCB > SCB   o. FCB = SCB   p. SNM < SCB   q. SNM > SCB   r. SNM = SCB

s. TD < SCB   t. TD > SCB   u. TD = SCB

* Based on negative ranks.

¶ Based on positive ranks.

3.3 Effects of age, gender, personality traits perspective taking and empathy

We investigated whether the positive impact of "predefined question stickers", "thinking face reaction", "private commenting" and the negative impact of "sentence openers" regarding perceived persuasiveness on willingness to challenge compared to the standard comment box were impacted by age, gender, Big Five personality traits, perspective-taking and empathy. We computed a difference score for each user and each interface by subtracting the level of the standard comment box reflected in users' ratings from the level of influence of presented interfaces. The difference score ranged from -6 to 6. Binomial logistic regression was performed using this score as a dependent variable to ascertain the effects of age, gender, personality traits, perspective taking and empathy on the likelihood that participants consider the presented interfaces to be more or less persuasive than the standard comment box. Positive difference scores were encoded as 1 which indicates that users found the presented interfaces more persuasive than the standard comment box. Any scores of zero or below were encoded as 0, which indicates either that there was no difference between the standard comment box and the interfaces or that users rated interfaces as less persuasive than the standard comment box.

A binomial logistic regression was performed to ascertain the effects of age, gender, personality traits, perspective taking and empathy on persuasiveness ratings. Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box-Tidwell [65] procedure. A Bonferroni correction was applied using all eight terms in the model resulting in statistical significance being accepted when $p < 0.0027$. Based on this assessment, all continuous independent variables were found to be linearly related to the logit of the dependent variable. The Hosmer and Lemeshow's tests suggest that the models for PQS, TFR, PC and SO provide good fit to the data. The result of the binary logistic regression estimates for each interface are presented in Table 3. Age and perspective taking were significantly positively associated with "predefined question stickers" ($b = 0.04$, $p < .01$ and $b = 0.08$, $p < .05$ respectively) such that increasing age and perspective-taking were associated with increased likelihood of positive influence of "predefined question stickers" relative to the standard comment box. However, increasing openness to experience was associated with a reduction in the likelihood of positive influence ($b = - 0.21$, $p < .01$). Openness to experience was associated with belief in the persuasiveness of the "thinking face reaction" such that as it increased, the positive influence of "thinking face reaction" of 1.24. ($b = - 0.24$, $p < .01$). Age was significantly negatively associated with "private commenting," such that as age increased, belief in the persuasiveness of private commenting decreased ($b = - 0.0 4$, $p < .01$)

Table 3. Binomial logistic regressions predicting the difference score between standard comment box and presented interfaces

Coefficients	PQS			TFR		
	B	SE	Exp(B)	B	SE	Exp(B)
Age	0.04*	0.01	1.04	-0.01	0.01	0.99
Gender (Female)	-0.38	0.31	0.68	-0.45	0.3	0.64
Extraversion	0.11	0.08	1.12	0.01	0.08	1.01
Agreeableness	-0.18	0.1	0.84	0.01	0.09	1.01
Conscientiousness	-0.12	0.1	0.89	0.06	0.1	1.07
Neuroticism	-0.02	0.08	0.98	0.03	0.08	1.03
Openness to experience	-0.21*	0.08	0.81	-0.24*	0.08	0.79
Empathy	-0.03	0.04	0.97	0.01	0.04	1.01
Perspective Taking	0.08**	0.04	1.08	-0.01	0.04	0.99
Constant	1.14	1.54	3.14	1.75	1.53	5.75
Modal summary						
Hosmer and Lemeshow X^2	5.28			9.87		
df	8			8		
p Value	.726			.27		
Nagelkerke R2	0.11			0.07		

*p < .01. **p < .05.

Table 3. Continued

Coefficients	PC			SO		
	B	S.E.	Exp(B)	B	S.E.	Exp(B)
Age	-0.04*	0.01	0.96	0	0.01	1
Gender (Female)	-0.15	0.31	0.86	-0.5	0.32	0.61
Extraversion	0.06	0.08	1.06	-0.06	0.09	0.94
Agreeableness	0.06	0.1	1.07	0.02	0.1	1.02
Conscientiousness	-0.09	0.1	0.91	-0.13	0.1	0.88
Neuroticism	0.01	0.08	1.01	-0.04	0.08	0.96
Openness to experience	-0.09	0.08	0.91	-0.14	0.08	0.87
Empathy	0.04	0.04	1.04	0.02	0.05	1.02
Perspective Taking	0	0.04	1	0.04	0.04	1.04
Constant	2.06	1.6	7.83	0.66	1.62	1.93
Modal summary						
Hosmer and Lemeshow X^2	9.22			4.83		
df	8			8		
p Value	.32			.77		
Nagelkerke R2	0.1			0.05		

*p < .01. **p < .05.

4   Discussion

4.1   Lack of tools for challenging misinformation on social media

While social media platforms such as Facebook rely on persuasive tactics to increase engagement with the platform which is important for commercial success [66], they do not seem to encourage critical thinking or enabling questioning of the content. On the contrary, several features they provide such as "hide post" or "unfriend" are used as tactics to discourage users from voicing concerns through facilitating less confrontational, yet less constructive, alternatives [67] as users may more easily avoid, rather than confront content. Our results showed that the majority of participants thought social media did not provide enough tools or ways to challenge misinformation. Although some features such as "commenting" or "direct messaging" may be used as tools to challenge misinformation, it can be argued that users do not regard them as tools provided to challenge misinformation. This necessitates the development of more persuasive or engaging tools, and ones which users would also perceive to be explicitly available for such purposes.

In our study, age was significantly negatively associated with reported lack of tools, such older participants were less likely to believe that social media offered tools for challenging misinformation. This might be because younger people, who were raised in an environment with a wealth of digital possibilities are more receptive to different social media features than older people [68]. This difference is to be expected given that young people were raised in an environment with a wealth of digital possibilities, whereas older people learned about such digital opportunities as adults [69] This may also explain why older adults are less likely to report correcting others [70], though we recognise that there are other potential explanations here, such as different perceptions of acceptable behaviours (see below).

4.2   Influence of social media interfaces on challenging misinformation

Social media interfaces may be considered as far more than just a means of providing access to information. They are also a way to alter users' attitudes and behaviours by creating opportunities for persuasive interaction [17]. Previous research investigated persuasive technologies to motivate users to change behaviours relating to themselves such as diet [71] and physical activity [72]. However, as challenging misinformation also includes interpersonal relationships in the online environment, the results may shed light on the influence of these persuasive techniques in interactive processes. Our results showed that "predefined question stickers" (suggestion), "thinking face reaction" (liking) and "private commenting" (reduction) influenced users the most and "sentence openers" (tunneling) influenced users the least to challenge misinformation. The suggestion technique was perceived as more persuasive than the standard comment box. In other words, people prefer the system to offer them prepared options to facilitate challenging misinformation, rather than having to write comments themselves. Along with providing suggestions to users, software-imposed interaction structures also make communication more impersonal [73], which may affect users' perceptions of information.

The effectiveness of the "thinking face reaction" shows that liking as a technique positively impacted users' willingness to challenge misinformation relative to the standard comment box. In addition to being visually appealing, the thinking face also allows the user to provide an impersonal response by clicking a reaction, which may account for the positive influence. Our findings showed that people found "private commenting", which is a preferred way to challenge misinformation [42], more persuasive than the standard comment box. In addition to simplifying the process, doing so in a private way may also contribute to its positive impact. As much as a reduction technique implemented in the form of private commenting seems to be successful, the inability of others to view the interaction is a concern as research shows that observing corrections helps individuals change their own misperceptions [74].

Tunneling in the form of "sentence openers" was found to be less persuasive than the standard comment box. This might be because although participants are given the starters of the sentences, they still need to write their arguments which may require more effort on their side. Prior research identified the importance of perceived ease of use in users' acceptance of information technologies [75]. In this case, solely guiding the user does not seem to be an effective strategy to motivate users to challenge misinformation. However, future work could investigate whether it helps users create more constructive responses.

4.3 Effect of age, openness to experience, and perspective taking on influence of the interfaces on challenging misinformation

Age, openness to experience, and perspective taking are moderated the effect of the interfaces that persuaded people more than the standard comment box. The positive association between age and "predefined question stickers" indicated that older adults found using prewritten statements more influential than the standard comment box. The negative association between age and "private commenting" shows that younger adults found it more influential to challenge misinformation in private. These findings suggest that as individuals age, concerns about self-presentation reduce and they feel more comfortable expressing themselves in public in line with previous research indicating that older adults tend to be less self-conscious and report fewer experiences of negative feelings such as shame, guilt and embarrassment relative to younger adults [76]. However, as people age, they choose to challenge misinformation in a more impersonal way, which may be an indication that they value their relationships with others. Perspective-taking is another predictor for "predefined question stickers". We found that users who have a greater tendency to consider events from the viewpoints, feelings, and reactions of others [77] find these stickers more persuasive than the standard comment box. As perspective-taking is related to make the same attributions to the target [31] , it could be argued that participants are more likely to accept these stickers if they receive them as a response. Openness to experience negatively predicted both "predefined question stickers" and "thinking face reaction," such that users with higher levels of openness to experience did not feel that these two interfaces were more persuasive than the standard comment box. Individuals high in openness to experience, which refers to having an active imagination or artistic interests [62] may possess unique cognitive processes and

thinking styles that enable them to challenge misinformation effectively in their own ways or using innovative methods. Therefore we anticipate they may not need any additional tools to challenge misinformation as they might already utilise their own methods of doing so.

4.4   Threats to validity

This study had several threats to validity that could affect the quality and generalisability of our findings. Our sample consisted of users from the U.K only. Research has shown that in Western societies open discussions and direct confrontation are more socially acceptable [78, 79]. The measure we used to assess the influence of the interfaces on willingness to challenge misinformation was based on self-report meaning potential biases such as social desirability bias.

We chose Facebook as an example of a SNS to test our proposed persuasive techniques, but it is possible that the results may differ for other SNS platforms. User cultures, interface design and user experience may differ across social media platforms [80, 81].

We acknowledge that our explanation of the results reflect one possibility. Alternative explanations can also be plausible. Therefore, future research is required to fully understand the results.

5   Conclusion and Future Work

We conducted a survey with 250 participants to assess how users perceived the persuasiveness of seven PSD-informed design interfaces compared to the standard comment box on willingness to challenge misinformation. Our study provides novel contributions to Persuasive Technology by identifying which interfaces users consider to be most persuasive in motivating individuals to challenge others on social media. In addition, these results provide insights into techniques that can be utilised to persuade people who hesitate to confront the perpetrators in instances of racism, sexism or prejudice on social media. By demonstrating that particular techniques are perceived as persuasive in users' willingness to challenge we pave the way for future social network design features to motivate users to challenge when they encounter misinformation. We also noted that users' demographics and psychological factors impacted their evaluation of different techniques, with younger users tending to favour private messages more than their older counterparts.

One of the study limitations was the use of self-reporting to assess the influence of each interface on the willingness to challenge, though further work could experimentally assess changes in users' levels of challenge with varying design implementation. Similarly, although we selected just seven techniques from the PSD framework, future research could investigate whether other techniques influence willingness to challenge misinformation. Another area for future work would involve examining whether other variables such as self-efficacy or self-enhancement are associated with the use of PSD-informed techniques. Our results provide information regarding the level of influence

of each technique, but future research may further explore the reasons, or at least require users to provide a rationale, for the ratings they provided, and factors influencing their acceptance of persuasive techniques. Many people who witness acts of racism or prejudice refrain from confronting the offenders [82]. Future research could investigate whether these techniques might influence willingness to confront such perpetrators.

References

1. Allcott, H., Gentzkow, M.: Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives 31, 211-236 (2017)

2. Allcott, H., Gentzkow, M., Yu, C.: Trends in the diffusion of misinformation on social media. Research & Politics 6, 2053168019848554 (2019)

3. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y.: Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Transactions on Intelligent Systems & Technology 10, 1-42 (2019)

4. Bode, L.: User correction as a tool in the battle against social media misinformation. Georgetown Law Technology Review 4, 367 (2019)

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

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

7. Bode, L., Vraga, E.K.: See something, say something: correction of global health misinformation on social media. Health communication 1131 (2018)

8. Walter, N., Brooks, J.J., Saucier, C.J., Suresh, S.: Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. Health communication 36, 1776-1784 (2021)

9. Chadwick, A., Vaccari, C.: News sharing on UK social media: Misinformation, disinformation, and correction. (2019)

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

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

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

13. Vraga, E.K., Bode, L.: Using expert sources to correct health misinformation in social media. Science Communication 39, 621-645 (2017)

14. Bautista, J.R., Zhang, Y., Gwizdka, J.: US physicians' and nurses' motivations, barriers, and recommendations for correcting health misinformation on social media: Qualitative interview study. JMIR Public Health and Surveillance 7, e27715 (2021)

15. Gurgun, S., Arden-Close, E., Phalp, K.T., Ali, R.: Online silence: why do people not challenge others when posting misinformation? Internet Research (2022)

16. Oinas-Kukkonen, H.: A foundation for the study of behavior change support systems. Personal and Ubiquitous Computing 17, 1223-1235 (2013)

17. Oinas-Kukkonen, H., Harjumaa, M.: Towards deeper understanding of persuasion in software and information systems. In: ACHI '08: Proceedings of the First International Conference on Advances in Computer-Human Interaction, pp. 200-205. IEEE, (2008)

18. Fogg, B.J.: Creating persuasive technologies: an eight-step design process. In: Proceedings of the 4th International Conference on Persuasive Technology, pp. 1-6.

19. Langrial, S., Lehto, T., Oinas-Kukkonen, H., Harjumaa, M., Karppinen, P.: Native mobile applications for personal well-being: a persuasive systems design evaluation. 16th Pacific Asia Conference On Information Systems, (2012)

20. Widyasari, Y.D.L., Nugroho, L.E., Permanasari, A.E.: Persuasive technology for enhanced learning behavior in higher education. International Journal of Educational Technology in Higher Education 16, 1-16 (2019)

21. Nkwo, M., Orji, R.: Personalized persuasion to promote positive work attitudes in public workplaces. In: UMAP'19 Adjunct: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 185-190. (2019)

22. Wiafe, I., Koranteng, F.N., Owusu, E., Ekpezu, A.O., Gyamfi, S.A.: Persuasive social features that promote knowledge sharing among tertiary students on social networking sites: An empirical study. Journal of Computer Assisted Learning 36, 636-645 (2020)

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

24. Flaxman, S., Goel, S., Rao, J.M.: Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opinion Quarterly 80, 298-320 (2016)

25. Pariser, E.: The Filter Bubble: What the Internet is Hiding from You. Penguin UK (2011)

26. Spohr, D.: Fake news and ideological polarization: Filter bubbles and selective exposure on social media. Business Information Review 34, 150-160 (2017)

27. Blau, I., Barak, A.: How do personality, synchronous media, and discussion topic affect participation? Journal of Educational Technology & Society 15, 12-24 (2012)

28. Alqahtani, F., Meier, S., Orji, R.: Personality-based approach for tailoring persuasive mental health applications. User Modeling and User-Adapted Interaction 32, 253-295 (2022)

29. Halko, S., Kientz, J.A.: Personality and persuasive technology: an exploratory study on health-promoting mobile applications. In: International Conference on Persuasive Technology, pp. 150-161. Springer, (2010)

30. Davis, M.H.: A multidimensional approach to individual differences in empathy. JSAS Catalogue of Selected Documents in Psychology 10, 85 (1980)

31. Davis, M.H., Conklin, L., Smith, A., Luce, C.: Effect of perspective taking on the cognitive representation of persons: a merging of self and other. Journal of Personality and Social Psychology 70, 713 (1996)

32. Wang, X.: To communicate or not to communicate: factors predicting passengers' intentions to ask a driver to stop text messaging while driving. Health communication 31, 617-625 (2016)

33. Baughan, A., Petelka, J., Yoo, C.J., Lo, J., Wang, S., Paramasivam, A., Zhou, A., Hiniker, A.: Someone is wrong on the internet: Having hard conversations in online spaces. Proceedings of the ACM on Human-Computer Interaction 5, 1-22 (2021)

34. Dennen, V.P.: Pedagogical lurking: Student engagement in non-posting discussion behavior. Computers in Human Behavior 24, 1624-1633 (2008)

35. Bishop, J.: Increasing participation in online communities: A framework for human–computer interaction. Computers in Human Behavior 23, 1881-1893 (2007)

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

37. Mascheroni, G., Murru, M.F.: "I can share politics but I don't discuss it": everyday practices of political talk on Facebook. Social Media+ Society 3, 2056305117747849 (2017)

38. Geiger, N., Swim, J.K.: Climate of silence: Pluralistic ignorance as a barrier to climate change discussion. Journal of Environmental Psychology 47, 79-90 (2016)

39. Statista, Q.: Number of monthly active Facebook users worldwide as of 4th quarter 2015. (2016)

40. https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/

41. Valenzuela, S., Kim, Y., Gil de Zúñiga, H.: Social networks that matter: Exploring the role of political discussion for online political participation. International Journal of Public Opinion Research 24, 163-184 (2012)

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

43. Toulmin, S.E.: The Uses of Argument. Cambridge University Press (2003)

44. Tom Tong, S., Walther, J.B.: Just say ''no thanks'': Romantic rejection in computer-mediated communication. Journal of Social and Personal Relationships 28, 488-506 (2011)

45. Oyebode, O., Alqahtani, F., Orji, R.: Exploring for Possible Effect of Persuasive Strategy Implementation Choices: Towards Tailoring Persuasive Technologies. In: International Conference on Persuasive Technology, pp. 145-163. Springer, (2022)

46. Halttu, K., Oinas-Kukkonen, H.: Need for Cognition Among Users of Self-Monitoring Systems for Physical Activity: Survey Study. JMIR Formative Research 5, e23968 (2021)

47. Bandura, A.: Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes 50, 248-287 (1991)

48. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Steering user behavior with badges. In: 22nd International Conference on World Wide Web
pp. 95-106.

49. Waardenburg, T., Winkel, R., Lamers, M.H.: Normative social influence in persuasive technology: intensity versus effectiveness. In: International Conference on Persuasive Technology, pp. 145-156. Springer, (2012)

50. Stibe, A., Oinas-Kukkonen, H., Lehto, T.: Exploring social influence on customer engagement: a pilot study on the effects of social learning, social comparison, and normative influence. In: 2013 46th Hawaii International Conference on System Sciences, pp. 2735-2744. IEEE, (2013)

51. Cialdini, R.B., Trost, M.R.: Social influence: Social norms, conformity and compliance. In: Gilbert, D.T., Fiske, S.T., Lindzey, G. (eds.) The Handbook of Social Psychology, pp. 151-192. McGraw-Hill, New York, NY (1998)

52. Lapinski, M.K., Rimal, R.N.: An explication of social norms. Communication Theory 15, 127-147 (2005)

53. Chung, A., Rimal, R.N.: Social norms: a review. Review of Communication Research (2016)

54. Perkins, H.W.: The emergence and evolution of the social norms approach to substance abuse prevention. The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors and Clinicians, pp. 3-17 (2003)

55. Wang, J.Z., Zhang, A.X., Karger, D.R.: Designing for Engaging with News using Moral Framing towards Bridging Ideological Divides. Proceedings of the ACM on Human-Computer Interaction 6, 1-23 (2022)

56. Albertson, B.P.: 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, 45-66 (2020)

57. Adler, M., Rougle, E.: Building Literacy Through Classroom Discussion: Research-Based Strategies for Developing Critical Readers and Thoughtful Writers in Middle School. Education Review (2007)

58. Herrmann, B., Gallo, J.R.: Facilitating Discussion of Theory and Practice in Education Seminars. Networks: An Online Journal for Teacher Research 20, 2 (2018)

59. https://www.facebook.com/brand/resources/facebookapp/reactions/

60. Yang, M., Ren, Y., Adomavicius, G.: Engagement by Design: An Empirical Study of the "Reactions" Feature on Facebook Business Pages. ACM Transactions on Computer-Human Interaction (TOCHI) 27, 1-35 (2020)

61. Cramer, H., De Juan, P., Tetreault, J.: Sender-intended functions of emojis in US messaging. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 504-509.

62. Rammstedt, B., John, O.P.: Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 41, 203-212 (2007)

63. Prot, S., Gentile, D.A., Anderson, C.A., Suzuki, K., Swing, E., Lim, K.M., Horiuchi, Y., Jelic, M., Krahé, B., Liuqing, W.: Long-term relations among prosocial-media use, empathy, and prosocial behavior. Psychological Science 25, 358-368 (2014)

64. Dovidio, J.F., Johnson, J.D., Gaertner, S.L., Pearson, A.R., Saguy, T., Ashburn-Nardo, L.: Empathy and intergroup relations. In: Mikulincer, M., Shaver, P.R. (eds.) Prosocial Motives, Emotions, and Behavior: The Better Angels of Our Nature. American Psychological Association (2010)

65. Box, G.E.P., Tidwell, P.W.: Transformation of the independent variables. Technometrics 4, 531-550 (1962)

66. Fogg, B.J., Iizawa, D.: Online persuasion in Facebook and Mixi: A cross-cultural comparison. In: International Conference on Persuasive Technology, pp. 35-46. Springer, (2008)

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

68. Arli, D.: Does social media matter? Investigating the effect of social media features on consumer attitudes. Journal of Promotion Management 23, 521-539 (2017)

69. Prensky, M.: Digital natives, digital immigrants. On the Horizon 9, (2001)

70. Bode, L., Vraga, E.K.: Correction experiences on social media during COVID-19. Social Media + Society 7, (2021)

71. Lee, M.K., Kiesler, S., Forlizzi, J.: Mining behavioral economics to design persuasive technology for healthy choices. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 325-334. (2011)

72. Matthews, J., Win, K.T., Oinas-Kukkonen, H., Freeman, M.: Persuasive technology in mobile applications promoting physical activity: a systematic review. Journal of Medical Systems 40, 1-13 (2016)

73. Walther, J.B.: Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Communication Research 23, 3-43 (1996)

74. Vraga, E.K., Bode, L.: I do not believe you: How providing a source corrects health misperceptions across social media platforms. Information, Communication & Society 21, 1337-1353 (2018)

75. Venkatesh, V.: Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. Information Systems Research 11, 342-365 (2000)

76. Henry, J.D., von Hippel, W., Nangle, M.R., Waters, M.: Age and the experience of strong self-conscious emotion. Aging & Mental Health 22, 497-502 (2018)

77. Galinsky, A.D., Moskowitz, G.B.: Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. Journal of Personality and Social Psychology 78, 708 (2000)

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

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

80. Weller, K.: Trying to understand social media users and usage: The forgotten features of social media platforms. Online Information Review 40, 256-264 (2016)

81. Dumbrell, D., Steele, R.: Social media technologies for achieving knowledge management amongst older adult communities. Procedia-Social and Behavioral Sciences 147, 229-236 (2014)

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