

# Northumbria Research Link

Citation: Craig, Michael and Vijaykumar, Santosh (2023) One dose is not enough: the beneficial effect of corrective COVID-19 information is diminished if followed by misinformation. *Social Media + Society*, 9 (2). p. 205630512311612. ISSN 2056-3051

Published by: SAGE

URL: <https://doi.org/10.1177/20563051231161298>  
<<https://doi.org/10.1177/20563051231161298>>

This version was downloaded from Northumbria Research Link:  
<https://nrl.northumbria.ac.uk/id/eprint/51446/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

**One dose is not enough: the beneficial effect of corrective COVID-19 information  
is diminished if followed by misinformation**

**Abstract**

The World Health Organization (WHO) released a series of mythbuster infographics to combat misinformation during the COVID-19 infodemic. While the corrective effects of such debunking interventions have typically been examined in the immediate aftermath of intervention delivery; the durability of these corrective effects and their resilience against subsequent misinformation remains poorly understood. To this end, we asked younger and older adults to rate the truthfulness and credibility of ten statements containing misinformation about common COVID-19 myths, as well as their willingness to share the statements through social media. They did this three times, before and after experimental interventions within a single study session. In keeping with established findings, exposure to the WHO's myth-busting infographics - (i) improved participants' ratings of the misinformation statements as untruthful and uncredible and (ii) reduced their reported willingness to share the statements. However, within-subject data revealed these beneficial effects were diminished if corrective information was presented shortly by misinformation, but the effects remained when further corrective information was presented. Throughout the study, younger adults rated the misinformation statements as more truthful and credible and were more willing to share them. Our data reveal that the benefit of COVID-19 debunking interventions may be short-lived if followed shortly by misinformation. Still, the effect can be maintained in the presence of further corrective information. These outcomes provide insights into the effectiveness and durability of corrective information and can influence strategies for tackling health-related misinformation, especially in younger adults.

**26 Introduction**

27 Public health agencies have launched digital communication interventions to address  
28 misperceptions seeded by the online circulation of COVID-19 misinformation. The severity of the  
29 COVID-19 misinformation problem is reflected in the World Health Organisation (WHO) labelling  
30 it an “infodemic” (Calleja et al., 2021). An integral part of combative strategies is the dissemination  
31 of ‘corrective information’, which involves debunking misleading claims circulating on social media  
32 (Bavel et al., 2020). A classic example is the “Mythbusters” intervention by the WHO, a digital  
33 resource where infographics are used to address public misperceptions related to a range of COVID-  
34 19 misinformation (World Health Organization, 2022). Recent work shows that beliefs in COVID-  
35 19 misinformation may be reduced through a single exposure to corrective information (Vijaykumar  
36 et al., 2021; Vraga & Bode, 2021). Randomised controlled trials of brief 60-second exposure to  
37 corrective infographics have yielded minor positive effects supporting arguments about the  
38 scalability of such nimble interventions (Agle et al., 2021). However, how long does the protective  
39 effect of a single dose last? What happens if people are exposed to misinformation shortly after a  
40 dose of corrective information? Misinformation research indicates that light-touch interventions  
41 (such as single corrections, infographics or 'accuracy nudges') dissipate swiftly, even over a duration  
42 of seconds in some cases (Roozenbeek et al., 2021). Thus, comprehending the underpinnings of  
43 corrective effects and factors that drive their durability has major implications for implementing fact-  
44 checks/accuracy nudges and other light-touch interventions in social media environments.

45 While studies examining the durability of corrective debunking interventions suggest a finite  
46 benefit, prebunking interventions that seek to inoculate audiences before misinformation have shown  
47 to confer a longer-lasting effect (two to six weeks) (Linden et al., 2021). Prebunking might be ideal  
48 for inoculating the public against misinformation in a general sense, but black swan events like the  
49 COVID-19 or even other infectious disease outbreaks like Ebola and Zika arrive under atypical  
50 conditions. Specifically, these pertain to unique disease characteristics, minimal understanding of the

51 nature of their impact on human health, and mystery surrounding modes of transmission, all of which  
52 create a fertile breeding ground for misinformation to emerge and proliferate. New misinformation  
53 content specific to these conditions then emerge and spread, commanding public health agencies to  
54 respond swiftly using debunking strategies. Research on debunking political misinformation has  
55 demonstrated that the effects of reaffirming truths and retracting falsehoods resulted in participants  
56 re-believing the misinformation after a week, suggesting a “continued influence” of misinformation  
57 (Swire et al., 2017).

58         Moreover, the endurance of post-information corrective effects may be strengthened by  
59 repeated exposure to corrective information through strategies like booster sessions and weakened by  
60 decaying factors like political predispositions and pre-existing attitudes (Carnahan et al., 2021).  
61 Understanding the specific mechanisms underpinning these findings allows the development of  
62 targeted interventions to reduce misinformation effects. These problems have been investigated less  
63 in the public health context, with the COVID-19 pandemic amplifying the need for more research to  
64 understand effective debunking strategies.

65         To achieve this, three primary gaps in our understanding of the durability of corrective  
66 information must be addressed. The first involves assessing the durability of the impact of real-world  
67 public health communication interventions like the WHO’s infographics. Second, durability  
68 assessments need to incorporate the ephemeral and transient nature of the flow of information on  
69 social media where users could be exposed to a trove of information, often with competing narratives  
70 within minutes. Third, the seemingly changeable impact of age on the durability of corrective effects  
71 must be understood. We first discuss the cognitive and behavioural outcomes that corrective  
72 information interventions seek to influence and then provide a rationale for focusing on age as a  
73 critical individual factor in this process.

74  
75 **Cognitive and Behavioural Impacts of Corrective Information**

76 Our evaluation of the durability of corrective information interventions like the Mythbusters  
77 is premised on its ability to steer and sustain three cognitive and behavioural responses in the desired  
78 direction: perceived truthfulness, perceived credibility, and intention-to-share the information.

79

80 *Perceived Truthfulness:* Debunking interventions using corrective information are commonly  
81 evaluated based on their ability to shift audience's beliefs away from misinformation and strengthen  
82 their ability to correctly identify the accuracy of these messages. Evaluating the accuracy of the  
83 content becomes especially important while engaging with the social media ecosystem where  
84 audiences could be exposed to information of various levels or provenance, or "shades of truth" from  
85 fully false to partly false and fully true (Lockyer et al., 2021; Wang, 2017). Partly false content can  
86 be especially problematic given that it can entrench beliefs in misinformation and undermine the  
87 effectiveness of corrective information (Freelon & Wells, 2020; Southwell et al., 2018). Low levels  
88 of knowledge, dependence on heuristic cues like fluency, and reasoning ability can affect the ability  
89 to discern between accurate and inaccurate information (Pennycook & Rand, 2021), but the role of  
90 repeated exposure to messages is especially important. The illusory truth effect says that people tend  
91 to perceive information as truer if they have been exposed to it before (Hassan & Barber, 2021). This  
92 means, for instance, that being exposed to the same COVID-19 falsehood arriving via different  
93 WhatsApp groups or connections can enhance the truthfulness of misinformation. The criticality of  
94 timely dissemination of corrective information is amplified even further in such situations. While  
95 some uncertainty remained over relevance of the illusory truth effect in claims that are obviously true  
96 or false, recent evidence from a simulated experiment shows its influence persisted across  
97 ambiguous and unambiguous statements (Fazio et al., 2019). The magnitude of the effect of  
98 repetition in the context of a real-world public health intervention such as the WHO's Mythbusters is  
99 less understood and will be investigated in this study.

100

101 *Message Credibility:* Assessments about the accuracy of messages (perceived truthfulness), in turn  
102 are shown to affect perceptions about its' credibility (Jung et al., 2016). The perceived credibility of  
103 the message is defined as "an individual's judgment about the veracity of the content of the  
104 communication" (Appelman & Sundar, 2016). Four broad categories of factors that can influence the  
105 perceived credibility of corrective information, and its potential to persuade audiences away from  
106 believing misinformation (Lee & Shin, 2021a). 1) Message Characteristics: Messages that are  
107 consistent, as opposed to discordant, with one's beliefs systems might seem more credible because  
108 these are easier to recall and can be used to arrive at a conclusion (Nickerson, 1998; Zhou & Shen,  
109 2022). While evidentiary devices like statistics, graphs and quotes are often included to strengthen  
110 the credibility of corrective information, the 'truth bias' imposed by these strategies can also be  
111 leveraged to spread misinformation (Newman et al., 2012). The frequency with which messages are  
112 disseminated could play a critical role in enhancing their perceived credibility, as suggested earlier  
113 by the 'illusory truth' bias. In other words, if repeated exposure to misinformation can enhance the  
114 believability of false claims, it is plausible that a similar strategy could be used with corrective  
115 information for beneficial effects. However, corrective information by public health agencies like the  
116 WHO's mythbusters are often online resources in stasis on their website with no possible  
117 determination about how frequently audiences are exposed to them. One of the focal points of this  
118 study is to determine if a single exposure can bear lasting effects. 2) Source Characteristics: Specific  
119 attributes of information sources have proved useful in strengthening to benefits of corrective  
120 interventions as they provide important social cues(Ecker et al., 2022) . For instance, corrective  
121 interventions delivered by government authorities and health experts minimise misinformation belief  
122 to a greater extent than social peers (van der Meer & Jin, 2020). Messages seem truer when delivered  
123 by credible, as opposed to non-credible sources, or sources who seem familiar, attractive and  
124 powerful (Briñol & Petty, 2009; Nadarevic et al., 2020). However, people's inattentiveness and  
125 forgetfulness could undermine source effects on credibility judgments with some studies showing

126 that people can discern the veracity of (mis)information irrespective of the source (Vijaykumar et al.  
127 2021). Based on this evidence, our experimental stimuli mention the source of the mythbusters  
128 (WHO) but measures the perceived credibility of the message as opposed to the institution. 3)  
129 Channel: Channel considerations pertain to the modality (images vs. text), synchronicity (delivered  
130 in real time vs. delivered with a delay), and medium (traditional media vs. social media) (E.-J. Lee &  
131 Shin, 2021b). Of most relevance to this study is consistent evidence that images possess greater  
132 persuasive power than simply text and are perceived to be more informative and useful (Lee & Shin,  
133 2021b; Lee et al., 2022a). Building on this strand, mythbuster infographics disseminated by the  
134 WHO consistently minimised COVID-19 misperceptions (Vijaykumar et al., 2021; Vraga & Bode,  
135 2021). 4) Individual factors: While several individual characteristics such as knowledge and  
136 numerical literacy render individuals vulnerable to misbelieving misinformation to be credible  
137 (Roozenbeek et al., 2020a), our study seeks to shed further clarity on the inconclusive debates around  
138 the role of age. Our arguments are presented at the end of this sub-section.

139 *Intention-to-Share:* In COVID-19 and non-COVID-19 contexts, evidence shows that  
140 messages that are perceived to be credible are also more likely to be shared (Song et al., 2023;  
141 Stefanone et al., 2019). Sharing behaviour underpins the extent to which information  
142 spreads or goes “viral” on social media, potentially influencing behavioural intentions (Alhabash &  
143 McAlister, 2015). Viral content can quickly reach and influence greater numbers of audiences, with  
144 dangerous or beneficial effects shared more widely and quickly depending on the nature of the  
145 content. For instance, health misinformation about COVID-19 vaccines that goes viral on social  
146 media can infuse doubts about the side effects of the vaccine leading to vaccine hesitancy and  
147 potentially vaccine refusal (Dror et al., 2020). While sharing accurate information potentially confers  
148 greater societal benefits, research has shown that misinformation is shared more widely and quickly  
149 possibly because of its novelty and ability to elicit emotional reactions (Vosoughi et al., 2018).  
150 Among health communication strategies that can trigger further dissemination by audiences, recent

151 research shows that infographics trigger greater sharing intentions especially while messaging about  
152 health issues related to proximal health behaviours or outcomes (e.g., a flu shot) (incomplete) and  
153 can thus be especially relevant during infectious disease outbreaks (Lee et al., 2022). Previous work  
154 has also demonstrated that that the WHO's mythbusters infographics can positively affect sharing  
155 intentions related to accurate misinformation (Vijaykumar et al., 2021). We build on this investigate  
156 how sharing intentions fluctuate in the face of repeated exposure to misinformation or corrective  
157 information.

158

### 159 **Age and Misinformation**

160 Of the various individual level factors that drive vulnerability to misinformation, the evidence  
161 surrounding the relationship between age and misinformation commands is particularly conflicting.  
162 For instance, older adults (over 65 years of age) were seven times more likely to share political fake  
163 news as opposed to younger adults aged 18-29 (Guess et al., 2019a). These findings are explained by  
164 lower levels of digital media literacy among older adults and the detrimental effect of age-related  
165 memory decline on increased susceptibility to the 'illusion of truth' effect (where repeated exposure  
166 to a false claim can make it seem like the truth). Similar explanations have been provided for  
167 findings which suggest that older white men are more likely to be engaged with fake news sources  
168 (Grinberg et al., 2019a). Analyses of media consumption patterns show that greater television  
169 consumption by older adults (55+) might expose them to ordinary bias and agenda setting by the  
170 mainstream media (Allen et al., 2020a). The dependence on information they are familiar with  
171 (fluency), challenges with source recall and difficulties with detecting deception are other reasons  
172 why older people may be vulnerable to misinformation (Brashier & Schacter, 2020).

173 However, other studies have found weak associations between older age and susceptibility to  
174 COVID-19 misinformation in four of five countries (the only exception being Mexico) (Rozenbeek  
175 et al., 2020a). A randomized online survey experiment of the effectiveness of the WHO's mythbuster

176 infographics found that younger adults (18-35) demonstrated stronger beliefs in misinformation than  
177 participants 55 years or older (Vijaykumar et al., 2021). These findings are partly explained by the  
178 ability of older adults to accumulate facts over time and evaluate the veracity of new information  
179 based on how it aligns with their general knowledge (Brashier & Schacter, 2020). An experiment  
180 testing the illusory truth effect between younger and older adults finding minimal differences  
181 between the two groups (Mutter et al., 1995; Parks & Toth, 2006). In sum, the evidence around the  
182 effect of age on vulnerability to misinformation is mixed with divergent findings across political  
183 misinformation, health misinformation and more generic misinformation like trivia.

184 **Study Aims & Hypotheses:** To this end, we asked younger and older adults to rate the truthfulness  
185 and credibility of ten statements containing misinformation about different COVID-19 myths, as  
186 well as their willingness to share the statements through social media. They did this on three  
187 occasions within a single session: (i) on entering the study (Baseline), (ii) following exposure to ten  
188 corrective infographics developed by the WHO, one per misinformation statement (Intervention 1),  
189 and then (iii) after exposure to ten WhatsApp messages (Intervention 2). Five of the WhatsApp  
190 messages contained *misinformation* relating to five of the statements, and the remaining five  
191 contained *corrective information* relating to the other five statements.

192 In keeping with existing literature, we predicted that exposure to the debunking infographics  
193 (Intervention 1) would improve participants' ratings of the misinformation statements as untruthful  
194 and unbelievable and reduce their willingness to share the statements through social media. Critical to  
195 the current study, should the benefit of corrective information be abated by subsequent  
196 misinformation, we hypothesised the effect of Intervention 1 should be reduced, at least somewhat,  
197 in response to Intervention 2, but *only* for the five statements that receive WhatsApp messages  
198 containing misinformation. For the five statements that received a second 'dose' of corrective  
199 information in Intervention 2, we predicted that the benefit of Intervention 1 should be maintained,  
200 and possibly improved, should two 'doses' - in proximity - be better than one. Finally, if older adults

201 are less susceptible to COVID-19 misinformation, they should correctly rate the misinformation  
202 statements as less truthful and credible and be less willing to share them. Because of this,  
203 intervention effects may be less pronounced in this population  
204

## 205 **Methods**

### 206 **Participants**

207 An a priori analysis of the sample required was conducted using G\*Power (Version 3.1.9.7).  
208 To detect a difference between age groups with a medium effect size ( $d = 0.50$ ), 0.05 probability of  
209 error, and 0.90 power, a total sample of 172 participants were required ( $n = 86$  per age group). We  
210 exceeded this target through the recruitment of 231 younger adults aged 18-35 years old (43 males,  
211 186 females, 2 other;  $M$  age = 25.44 years,  $SD = 5.13$  years; age range = 18-35 years) and 237 older  
212 adults aged 55 years old and above (112 males, 125 females;  $M$  age = 62.54 years,  $SD = 6.12$  years;  
213 age range = 55-81 years). Categorisation of younger and older adults as those aged 18-35 years and  
214 55+ years old, respectively, was based on commonly used age ranges in psychological and  
215 biomedical literature (cite). These individuals were recruited through Qualtrics' panel of survey  
216 respondents. In addition, participants were required to fit our criteria for younger and older adults  
217 (see above), live in the United Kingdom, and be WhatsApp users aware of COVID-19. Aside from  
218 age (younger vs. older), we had no a priori predictions surrounding the contribution of other  
219 demographic factors, for example, gender and employment status, and thus did not control for these  
220 factors in our recruited sample. Figure 1 provides a visual overview of participant demographics,  
221 which were broadly representative of the general population. Data collection commenced on  
222 December 15<sup>th</sup> 2020 and culminated on March 10<sup>th</sup> 2021. Throughout this time, the United Kingdom  
223 remained under relatively severe "lockdown" restrictions, including mask-wearing, social distancing,  
224 and restricted mixing of households. All participants provided written informed consent to  
225 participating in the study before responding to the survey questions. Given the nature of the study,

226 when being debriefed, participants were directed towards truthful COVID-19 information about the  
227 topics covered in the study. The study was approved by the Faculty Ethics Committee at a large  
228 university in England (Ref: 120.1520).

229  
230 <<INSERT FIG 1 ABOUT HERE>>

231  
232 **Fig 1. Participant demographics.** The figure summarises participant demographic information for  
233 our younger (n = 231) and older (n = 237) adult groups. Details are shown regarding participants'  
234 ages, gender, geographic location in the United Kingdom, highest education level, current  
235 employment status, and annual salary.

236  
237 **Design**

238 To examine whether the beneficial effect of corrective COVID-19 interventions can  
239 withstand subsequent misinformation in younger and older adults, we employed a repeated measures  
240 design with between-subject factor *age group* (younger adults vs older adults) and within-subject  
241 factors *time of test* (Baseline vs Intervention 1 vs Intervention 2) and *truthfulness of information*  
242 *presented in Intervention 2* (corrective information vs misinformation). The study took place in a  
243 single session and was delivered online through the research platform Qualtrics.

244  
245 **Materials**

246 From the WHO's COVID-19 myth-buster webpage, which offers corrective infographics to  
247 debunk prevalent COVID-19 misinformation online, we selected five themes: *therapeutics*,  
248 *environment*, *behaviour*, *foodstuffs*, and *transmission*. **Ten infographics** (two per theme) were  
249 selected from the WHO's website. Within the remit of the limited number of infographics available,  
250 the two infographics selected for each theme were matched as closely as possible on their topic and

251 content, for example, that experiencing *cold temperatures* and *hot temperatures* can cure COVID-19  
252 (*environment* theme). These infographics were presented in the Intervention 1 phase – see Procedure.

253 Based on these ten infographics, we developed corresponding **misinformation statements**.  
254 For example, for an infographic tackling the myth that garlic can cure COVID-19 (*foodstuffs* theme),  
255 the following statement was prepared: "*Garlic can cure me of the Coronavirus (COVID-19)*".  
256 Similarly, for an infographic tackling the myth that COVID-19 can be transmitted through 5G  
257 networks (*transmission* theme), the following statement was developed: "*Viruses like Coronavirus*  
258 (*COVID-19*) *can be spread through mobile networks like 5G*". These ten misinformation messages  
259 were presented to participants in each phase of our study. They were asked to rate the truthfulness  
260 and credibility of the statements and their willingness to share them through social media - see  
261 Procedure.

262 Further to the above, based on the ten misinformation statements and linked corrective  
263 infographics, we developed ten graphics designed in the form of **forwarded WhatsApp messages**.  
264 Each WhatsApp message related to one of the ten misinformation statements. Critical to the purpose  
265 of the current study, these messages contained either (i) corrective information (total = 5), or (ii)  
266 misinformation (total = 5). For each of the five themes of misinformation, one WhatsApp message  
267 (e.g., *hot temperatures cure COVID-19*) contained correct information, e.g., "*research shows that*  
268 *hot temperatures do not cure COVID-19*". The other WhatsApp message (e.g., *cold temperatures*  
269 *cure COVID-19*) contained misinformation, e.g., "*research shows that hot temperatures can cure*  
270 *COVID-19*". These graphics were presented to participants in the Intervention 2 phase – see  
271 Procedure. All materials are available through the project's OSF site: <https://osf.io/4qm7y/>

272 The choice of WhatsApp-based stimuli for this study was based on several reasons.  
273 WhatsApp is the most used messaging service in the UK with more than 40 million users and was  
274 one of the global vectors of misinformation during the COVID-19 pandemic (cite). Resultantly,  
275 several organizations including the World Health Organization and the International Fact Checking

276 Network launched WhatsApp-based interventions like tiplines to combat the spread and impact of  
277 misinformation.

## 278 **Measures**

279 To establish whether the beneficial effect of corrective COVID-19 information is resilient  
280 against exposure to subsequent misinformation, we employed three dependent variable measures  
281 concerning misinformation belief. These three measures were applied in each phase of our study:  
282 Baseline, Intervention 1 (corrective information), and Intervention 2 (correct information vs  
283 misinformation).

284 First, we applied a measure of perceived **truthfulness**, where participants are required to  
285 “rate the truthfulness” of information on a scale from 1 = “not at all” to 9 = “very”. This measure  
286 was based on methods investigating the perceived accuracy of health-related messages (Carey et al.,  
287 2020), and which was updated recently for the context of COVID-19 misinformation (Vijaykumar et  
288 al., 2021).

289 Second, a measure of message **credibility** was employed (Appelman & Sundar, 2016). This  
290 scale-based measure asks participants to rate how well (from 1 = *very poorly* to 9 = *very well*) three  
291 adjectives describe communication content: *accurate*, *authentic*, and *believable*. We amended the  
292 scale from a seven- to nine-point scale for the current study. Given that scale, reliability analyses  
293 suggest this three-item measure has high internal reliability ( $\alpha = 0.87$ ) (Appelman & Sundar, 2016),  
294 we averaged responses from the three sub-scores into a single score (min score = 1, max score = 9)  
295 for analyses. Chronbach’s analyses confirmed high internal reliability across the three scale items in  
296 the current study ( $\alpha > 0.9$  in all instances).

297 Third, given the importance of misinformation dissemination, a **sharing** measure was used to  
298 explore participants willingness to share messages containing misinformation through social media.  
299 Specifically, based on existing methods (C. S. Lee & Ma, 2012), participants are asked how likely  
300 they would *intend*, *expect*, and *plan* to share content through social media. A rating on a five-point

301 scale from 1 = *highly unlikely* to 5 = *highly likely* was collected for each verb. Chronbach's analyses  
302 confirmed high internal reliability across the three scale items in the current study ( $\alpha > 0.9$  in all  
303 instances). Because of this, we averaged responses from the three sub-scores into a single score (min  
304 score = 1, max score = 5) for analyses.

305

## 306 **Procedure**

307 Our experimental procedure was inspired by research investigating the correction of  
308 misinformation (Lewandowsky & van der Linden, 2021; Vijaykumar et al., 2021; Vraga & Bode,  
309 2021) and memory paradigms used to examine the effect of within-subject manipulations on memory  
310 accuracy during reconsolidation (Hupbach et al., 2007; Przybyslawski & Sara, 1997). Participants  
311 were informed that they were participating in a study investigating how we make judgements about  
312 COVID-19 information found online. The procedure comprised three phases and took place in a  
313 single session: Baseline, Intervention 1, and Intervention 2. During the **Baseline** phase, participants  
314 were presented sequentially ten misinformation messages relating to prevalent COVID-19 myths  
315 identified by the WHO (see Materials). For example, "*Garlic can cure me of the Coronavirus*  
316 (*COVID-19*)". For each statement, participants were asked to rate the truthfulness and credibility of  
317 the messages. Their willingness to share the messages through social media was also probed. There  
318 was no time limit to respond. These measurements provided a pre-intervention baseline for relative  
319 comparison to establish post-intervention effects.

320 In the subsequent **Intervention 1** phase, participants were presented corrective COVID-19  
321 information in the form of the WHO's COVID-19 myth-buster infographics (see Materials). Ten  
322 infographics were presented, one concerning each topic covered in the ten misinformation statements  
323 (e.g., garlic cures COVID-19). The infographics were presented sequentially and in a random order,  
324 each for 30 seconds (total duration = 5 minutes). This fixed duration ensured all participants received  
325 identical treatment and exposure to corrective stimuli, opposed to self-paced exposure as used in

326 related work (Basol et al., 2021). After exposure to the corrective information, participants rated the  
327 truthfulness and credibility of the same ten randomly ordered misinformation statements for a second  
328 time as presented in the Baseline phase. They were also again asked to rate their willingness to share  
329 the statements. We did this to establish whether, as in previous work, exposure to corrective  
330 information positively affects participants treatment of misinformation.

331         Following this, in the **Intervention 2** phase, participants were presented ten WhatsApp  
332 messages, each concerning one of the topics covered in the ten misinformation statements (see  
333 Materials). Critical to our hypotheses, five of the messages contained *misinformation* and five  
334 contained *corrective information*. This within-subject manipulation enabled us to examine whether  
335 the possible benefit of corrective information in the Intervention 1 phase is abated by subsequent  
336 misinformation. If so, a corrective effect from Intervention 1 should be reduced, at least somewhat,  
337 in response to Intervention 2, but only for the five statements that receive misinformation in the  
338 WhatsApp messages. For the reasons explained above, WhatsApp messages were ordered randomly  
339 and presented sequentially for 30 seconds (total duration = 5 minutes). After exposure to the  
340 WhatsApp messages, participants rated the truthfulness and credibility of the same ten randomly  
341 ordered misinformation statements presented in the Baseline and Intervention 1 phases for a third  
342 and final time. They were also again asked to rate their willingness to share the statements.

343

344

### 345 **Statistical Analyses**

346         For the Baseline, Intervention 1, and Intervention 2 phases, mean truthfulness, credibility,  
347 and sharing scores were computed for (i) the five COVID-19 topics that received corrective  
348 information in Intervention 2 and (ii) the five COVID-19 topics that received misinformation in  
349 Intervention 2. Data were analysed using SPSS (Version 26.0; IBM Corp, 2019). Truthfulness,  
350 credibility, and sharing measures were investigated using individual Repeated Measures ANOVAs

351 with between-subject factor age group (younger adults vs older adults) and within-subject factors  
352 time of test (Baseline vs Intervention 1 vs Intervention 2) and Intervention 2 manipulation (corrective  
353 information vs misinformation). Pairwise comparisons were used to examine within-subject changes  
354 in responses from one time point to another (effect of Intervention 1: Baseline vs Intervention 1).  
355 They were also used to compare – within each age group - mean scores for each study phase, e.g.,  
356 comparison of mean truthfulness scores recorded at the Intervention 2 phase for items that received  
357 corrective information in Intervention 2 vs Items that received misinformation in Intervention 2.  
358 Bonferroni corrections were applied to correct for multiple comparisons.

359

## 360 **Results**

### 361 **Perceived Truthfulness**

362 Figure 2A shows mean truthfulness scores for each study phase broken down by age group  
363 (younger vs older) and our Intervention 2 manipulation (corrective information vs misinformation).  
364 We observed a significant main effect of time of test ( $F(2,932) = 82.305, p < .001, \eta^2 = .150$ )  
365 because there was an improvement in ratings following the presentation of corrective information in  
366 Intervention 1 and worsening in response to Intervention 2, predominantly for items that received  
367 misinformation in this study phase. This was reinforced through a significant effect of our  
368 Intervention 2 manipulation ( $F(1,466) = 33.347, p < .001, \eta^2 = .109$ ), where those who received  
369 misinformation in Intervention 2 generally performed poorer than those who received corrective  
370 information in the same study phase. A significant interaction between time of test and our  
371 Intervention 2 manipulation was observed ( $F(2,932) = 38.358, p < .001, \eta^2 = .076$ ) because the  
372 effect of this intervention (corrective information vs misinformation) was largely restricted to the  
373 final phase of our study (see Figure 2A). Pairwise comparisons revealed that item subset scores did  
374 differ significantly during the Baseline phase (younger:  $t(230) = -2.046, p = .042$ ; older:  $t(236) = -$   
375  $3.374, p = .001$ ), but were matched following presentation of corrective infographics in Intervention

376 1 (younger:  $t(230) = 0.159, p = .874$ ; older:  $t(236) = -0.299, p = .765$ ). A negative change in scores  
377 for items that received misinformation in Intervention 2 resulted in a significant difference between  
378 item subset scores in this phase (younger:  $t(230) = -6.173, p < .001$ ; older:  $t(236) = -4.810, p < .001$ ).

379 Throughout the study, older adults outperformed younger adults in the truthfulness measure  
380 ( $F(1,466) = 87.732, p < .001, \eta^2 = .158$ ), where the former performed near ceiling. A significant  
381 interaction between time of test and age was observed ( $F(2,932) = 16.347, p < .001, \eta^2 = .027$ )  
382 because the effect of our Intervention 1 manipulation was more pronounced in younger adults.  
383 However, this was somewhat driven by near ceiling effects in older participants, i.e., there was little  
384 room for them to improve. There was no significant interaction between age and our Intervention 2  
385 manipulation ( $F(1,466) = 0.687, p = .408, \eta^2 = .001$ ), indicating that the effect of corrective  
386 information vs misinformation was comparable in younger and older adults. Furthermore, we found  
387 no three-way interaction between age, time of test, and Intervention 2 manipulation ( $F(2,932) =$   
388  $1.793, p = .167, \eta^2 = .004$ ). All significant effects from the RM ANOVA remained after controlling  
389 for multiple comparisons (Bonferroni corrected  $p$  value = .007).

390 When gender was included as a covariate in the RM ANOVA, no significant findings  
391 changed, and overall trends remained. We did however observe a significant main effect of gender  
392 ( $F(1,465) = 16.073, p < .001, \eta^2 = .033$ ) because males performed poorer in this measure. There  
393 were no two- or three-way interactions between gender and our other factors (all  $p > .112$ ), indicating  
394 that the effect of age, time, and intervention 2 manipulation were comparable across genders.

395  
396 <<INSERT FIG 2 ABOUT HERE>>

397  
398 **Fig 2. Performances in perceived truthfulness, message credibility, and sharing intention**  
399 **measures.** The line graphs show mean scores for the truthfulness, credibility, and sharing measures  
400 from each study phase broken down by between-subject factor age (younger vs older) and within-

401 subject factor Intervention 2 manipulation (corrective information vs misinformation). Blue lines  
402 show data from younger adults, and red lines show data from older adults. Solid lines refer to data  
403 for statements presenting truthful information in Intervention 2 (total = 5), and dashed lines refer to  
404 data for statements presenting novel misinformation in Intervention 2 (total = 5). In all cases, a lower  
405 score reflects superior performance. Error bars show the standard error of the mean. Post-hoc  
406 pairwise comparisons conducted individually for younger and older adults revealed significant  
407 declines in scores between intervention 1 and intervention 2 testing times for items that received  
408 corrective information in intervention 1 and misinformation in intervention 2 (all  $p < .005$ ).

409

### 410 **Message Credibility**

411 Figure 2B shows mean credibility scores for each study phase broken down by age group  
412 (younger vs older) and our Intervention 2 manipulation (corrective information vs misinformation).  
413 We observed a significant main effect of time of test ( $F(2,932) = 31.912, p < .001, \eta p^2 = .064$ )  
414 because there was an improvement in ratings following the presentation of corrective information in  
415 Intervention 1 and worsening in response to Intervention 2, predominantly for items that received  
416 misinformation in this study phase. This was reinforced through a significant effect of our  
417 Intervention 2 manipulation ( $F(1,466) = 119.552, p < .001, \eta p^2 = .204$ ), where those who received  
418 misinformation in Intervention 2 generally performed poorer than those who received corrective  
419 information in the same study phase. A significant interaction between time of test and our  
420 Intervention 2 manipulation was observed ( $F(2,932) = 26.744, p < .001, \eta p^2 = .054$ ) because the  
421 effect of this intervention (corrective information vs misinformation) was largely restricted to the  
422 final phase of our study (see Figure 2B). Pairwise comparisons revealed that item subset scores did  
423 differ significantly during the Baseline phase (younger:  $t(230) = -8.376, p < .001$ ; older:  $t(236) = -$   
424  $8.053, p < .001$ ), but were matched following presentation of corrective infographics in Intervention  
425 1 (younger:  $t(230) = -1.412, p = .159$ ; older:  $t(236) = -1.873, p = .062$ ). A negative change in scores

426 for items that received misinformation in Intervention 2 resulted in a significant difference between  
427 item subset scores in this phase (younger:  $t(230) = -5.180$ ,  $p < .001$ ; older:  $t(236) = -4.541$ ,  $p < .001$ ).

428 Throughout the study, older adults outperformed younger adults in the credibility measure  
429 ( $F(1,466) = 66.128$ ,  $p < .001$ ,  $\eta^2 = .124$ ), where the former performed near ceiling. A significant  
430 interaction between time of test and age was observed ( $F(2,932) = 7.544$ ,  $p < .001$ ,  $\eta^2 = .016$ )  
431 because the effect of our Intervention 1 manipulation was more pronounced in younger adults. There  
432 was no significant interaction between age and our Intervention 2 manipulation ( $F(1,466) = 0.860$ ,  $p$   
433  $= .835$ ,  $\eta^2 = .002$ ), indicating that the effect of corrective information vs misinformation was  
434 comparable in younger and older adults. Furthermore, we found no three-way interaction between  
435 age, time of test, and Intervention 2 manipulation ( $F(2,932) = 2.404$ ,  $p = .091$ ,  $\eta^2 = .005$ ). All  
436 significant effects from the RM ANOVA remained after controlling for multiple comparisons  
437 (Bonferroni corrected  $p$  value = .007).

438 When gender was included as a covariate in the RM ANOVA, no significant findings  
439 changed, and overall trends remained. We did however observe a significant main effect of gender  
440 ( $F(1,465) = 9.911$ ,  $p = .002$ ,  $\eta^2 = .021$ ) because males performed poorer in this measure. There was  
441 also a significant interaction between gender and our intervention 2 manipulation ( $F(1,465) = 8.597$ ,  
442  $p = .004$ ,  $\eta^2 = .018$ ) because the effect of our manipulation was more pronounced in males though,  
443 like the interaction between age and our intervention 2 manipulation, this was at least partially driven  
444 by females performing closer to ceiling and thus having less room for improvement. No other  
445 interactions were significant (all  $p > .300$ ).

#### 446 **Sharing Intention**

447 Figure 2C shows mean sharing scores for each study phase broken down by age group  
448 (younger vs older) and our Intervention 2 manipulation (corrective information vs misinformation).  
449 In keeping with our other measures, we observed a significant main effect of time of test ( $F(2,932) =$   
450  $16.330$ ,  $p < .001$ ,  $\eta^2 = .034$ ) because there was an improvement in ratings following the presentation

451 of corrective information in Intervention 1 and worsening in response to Intervention 2,  
452 predominantly for items that received misinformation in this study phase. This was reinforced  
453 through a significant effect of our Intervention 2 manipulation ( $F(1,466) = 47.706, p < .001, \eta^2 =$   
454  $.093$ ), where those who received misinformation in Intervention 2 generally performed poorer than  
455 those who received corrective information in the same phase. A significant interaction between time  
456 of test and our Intervention 2 manipulation was observed ( $F(2,932) = 6.752, p = .001, \eta^2 = .014$ )  
457 because the effect of this intervention (corrective information vs misinformation) was largely  
458 restricted to the final phase of our study (see Figure 2C). Pairwise comparisons revealed that item  
459 subset scores differed significantly during the Baseline phase (younger:  $t(230) = -5.241, p < .001$ ;  
460 older:  $t(236) = -3.376, p = .001$ ), Intervention 1 phase (younger:  $t(230) = -1.300, p = .195$ ; older:  
461  $t(236) = -2.625, p = .009$ ), and Intervention 2 phase (younger:  $t(230) = -2.768, p = .006$ ; older:  $t(236)$   
462  $= -3.389, p < .001$ ), though the magnitude of the difference was more pronounced following our  
463 Intervention 2 manipulation (see Figure 2C).

464 Throughout the study, older adults outperformed younger adults in the sharing intention  
465 measure ( $F(1,466) = 72.654, p < .001, \eta^2 = .135$ ), where the former performed near ceiling. A  
466 significant interaction between time of test and age was observed ( $F(2,932) = 7.745, p < .001, \eta^2 =$   
467  $.016$ ) because the effect of our Intervention 1 manipulation was more pronounced in younger adults.  
468 There was no significant interaction between age and our Intervention 2 manipulation ( $F(1,466) =$   
469  $1.202, p = .273, \eta^2 = .003$ ), indicating that the effect of corrective information vs misinformation  
470 was comparable in younger and older adults. We did find a three-way interaction between age, time  
471 of test, and Intervention 2 manipulation ( $F(2,932) = 3.889, p = .049, \eta^2 = .008$ ), but this effect did  
472 not survive correction for multiple comparisons (Bonferroni corrected  $p$  value =  $.007$ ). All other  
473 effects remained significant.

474 When gender was included as a covariate in the RM ANOVA, no significant findings  
475 changed, and overall trends remained. We did however observe a significant main effect of gender

476 (F(1,465) = 7.551,  $p = .006$ ,  $\eta^2 = .016$ ) because males performed poorer in this measure. There were  
477 no two- or three-way interactions between gender and our other factors (all  $p > .661$ ), indicating that  
478 the effect of age, time, and intervention 2 manipulation were comparable across genders.

479

480

## 481 **Discussion**

482 The durability of corrective information by public health agencies on misinformation beliefs  
483 among social media users has seldom been investigated. For example, Vraga and Bode (2021)  
484 investigated the efficacy of WHO's infographics similar to the stimuli used in our study but focused  
485 on placement and source and not on durability. Meanwhile, Basol and colleagues (2021) found that  
486 COVID-19 infographics were less effective than prebunking inoculation strategies to improve  
487 people's confidence in spotting misinformation and reduce their willingness to share it. However,  
488 they used UNESCO infographics which contained more generic educational content than specific,  
489 topic-specific debunking content in our stimuli.

490 In keeping with existing literature, we found that exposure to corrective information – the  
491 WHO's "Mythbuster" infographics – improved participants rating of misinformation statements as  
492 untruthful and uncredible. It also reduced their willingness to share the statement through social  
493 media. However, our data suggest this beneficial effect of a 'single dose' of corrective information is  
494 short-lived *if* it is followed shortly by exposure to misinformation (Intervention 2). Critically, this  
495 effect was observed only for items where misinformation was presented in Intervention 2: exposure  
496 to further corrective information (i.e., a 'double dose') did not result in *further* improvements. Still, it  
497 did maintain the benefit of a single dose of corrective information. These findings reveal that the  
498 lifespan of a single dose of corrective information may not be sufficient to deliver long-lasting  
499 protection against COVID-19 misinformation. Furthermore, outcomes may be of particular

500 importance for younger adults, who demonstrated higher misinformation belief and willingness to  
501 share throughout our study. We discuss these findings and possible explanations in turn.

502         The benefit of corrective information in the Intervention 1 phase resonates with established  
503 effects following the debunking of misinformation, including about COVID-19 (Kreps & Kriner,  
504 2020; Linden et al., 2021; Vijaykumar et al., 2021). In addition, this work has included observance of  
505 corrective effects following the WHO’s infographics application (Basol et al., 2021). Pinpointing the  
506 drivers of this positive change is difficult to establish in our design but might be explained  
507 straightforwardly through the influence of the information presented on attitudes towards  
508 misinformation. This explanation may also account for the diminished benefit seen in Intervention 2  
509 for the subset of statements for which misinformation was presented. Inherent differences between  
510 item subsets are unlikely to explain the within-subject effect of our Intervention 2 manipulation.  
511 Despite some initial differences between item subsets in the Baseline phase, the corrective effect of  
512 infographics in the Intervention 1 phase acted as a “leveller”: truthfulness, credibility, and sharing  
513 scores were well-matched when probed in the Intervention 1 phase, which immediately preceded our  
514 within-subject Intervention 2 manipulation. Nevertheless, to rule out the contribution of item-by-item  
515 effects, we acknowledge that it would be advantageous to replicate our findings using a set of  
516 statements that were closely matched in the Baseline phase. Still, it is important that irrespective of  
517 any differences between items, other than the effect of age group, all effects reported reflected  
518 within-subject changes that were in response to our experimental manipulations.

519         An alternative explanation for the observed effects is that our experimental design affected  
520 the content of retained memories pertaining to the common COVID-19 myths. This possibility is in  
521 keeping with evidence demonstrating that memories are not fixed and can be altered/updated (for  
522 better or worse) through exposure to subsequent information shortly following their initial  
523 acquisition and subsequent recall, which influence consolidation and reconsolidation processes,  
524 respectively (Dudai & Eisenberg, 2004; Loftus, 2005; Spiers & Bendor, 2014). Even subtle cues, less

525 prominent than used in the current study, are found to re-enter memories into a labile state (Hupbach  
526 et al., 2007). Such memory studies inspired our experimental design. Therefore, it is possible that the  
527 (mis)information presented in the current study updated existing traces, which was detected in  
528 subsequent questioning. Indeed, given that questioning often occurred several minutes post-  
529 intervention exposure, this suggests that the effects reported in the current study did not dissipate  
530 rapidly but remain *at least* over the time course of minutes. This duration may be further indicative  
531 of a contribution of memory to our findings. Our design does not allow us to confirm this but may  
532 offer inspiration for future work. Indeed, the contribution of memory mechanisms to misinformation  
533 is noted as a promising area of investigation (Linden et al., 2021).

534 Further to these possibilities, other factors may have contributed to our findings, and we  
535 cannot rule out the contribution of demand characteristics. But the likelihood of extensive influence  
536 of experimenter influence is low given that participants were (i) unaware of the exact purpose of the  
537 study, (ii) not informed whether presented information was truthful or not, and (iii) provided ratings  
538 of truthfulness, credibility, and sharing (in most cases) several minutes after exposure to  
539 (mis)information in intervention 1 and 2. Had stimuli exposure and ratings been collected  
540 simultaneously, this may be more likely. Therefore, we propose influence of presented information  
541 on attitudes, and possible contributions in memory, are more likely explanations.

542 It is of interest that there was no *extra* benefit in the second intervention phase for  
543 misinformation statements that received further corrective information. This might suggest that two  
544 ‘doses’ of corrective information within minutes of one another have no added benefit over a single  
545 dose. While this is possible, our data cannot account for differences in the strength of the effect that  
546 may influence its durability. Thus, while our study offers new insights into the limited and temporary  
547 effectiveness of a single dose of corrective information, we cannot make inferences about the  
548 durability of two doses, other than demonstrating no negative effect of a second dose, even when  
549 presented in a different medium to the first.

550           How can the striking effect of age in misinformation belief and willingness to share  
551 misinformation be explained? Heightened misinformation belief and willingness to share  
552 misinformation in younger adults is in keeping with recent findings, but data are mixed, and other  
553 misinformation research suggests an effect of age in the opposing direction (Allen et al., 2020;  
554 Grinberg et al., 2019; Guess et al., 2019; Roozenbeek et al., 2020; Vijaykumar et al., 2021). These  
555 findings may be influenced by a broad range of factors, including political ideology, religiosity, and  
556 social ideology, which we did not measure here but are known to contribute to misinformation belief  
557 (Grinberg et al., 2019; Swire et al., 2017). In addition, behaviours surrounding social media use may  
558 also have contributed. Specifically, greater use of social media platforms, particularly news-seeking  
559 behaviours (Edgerly, 2017), may have resulted in our younger adults being exposed to more  
560 corrective information, but also misinformation. Indeed, this population are reported to be more  
561 likely to see and share COVID-19 disinformation (Crime and Security Research Institute, 2020;  
562 Herrero-Diz et al., 2020; Ofcom, 2019) (Herrero-Diz et al., 2020; Crime and Security Research  
563 Institute, 2020; Ofcom, 2019).

564           The effects of our interventions were less pronounced in older individuals partially because  
565 they performed near ceiling and demonstrated very little belief in the misinformation statements.  
566 Additionally, older adults' may rely on their more extensive knowledge and critically evaluate new  
567 information (Umanath & Marsh, 2014). The same may hold in the current study. A further  
568 consideration is that sampling only WhatsApp users may have resulted in the recruitment of digital  
569 and media literate older adults who are experienced in fact-checking online. If so, our sample may  
570 not be truly representative of the older adult population. Ceiling effects meant a reduced capacity to  
571 observe a benefit of corrective information in our older sample. Thus, while the observed effects  
572 were more prominent in younger individuals, we cannot rule out that both age groups may have  
573 benefited equally from corrective information had our measures been more sensitive. Despite the  
574 age-related differences in scores and magnitude of our intervention effects, both age groups' levels

575 of belief in misinformation were relatively low in the current study. Intriguingly, our data show that  
576 even in cases of minimal misinformation belief, debunking strategies can be effective.

577         The last and possibly the most important finding from our study is the extent to which  
578 encountering misinformation after exposure to corrective misinformation diminishes the cognitive  
579 gains conferred by the latter. This finding is consistent with studies which discovered that strong  
580 misinformation messages “neutralised” the positive effects gained after exposure to communication  
581 about the consensus around climate change (Maertens et al., 2020). In the current social media  
582 context, these findings behove public health agencies to consider how the already fleeting impact of  
583 light-touch interventions such as Mythbusters might be further undercut by the very realistic prospect  
584 of subsequent exposure to misinformation. While it might be tempting to use these findings to call  
585 for corrective information to be delivered in a synchronised way between public health agencies and  
586 social media platforms, it is not clear how such strategies can be implemented on applications like  
587 WhatsApp where content is fully encrypted. These findings also call for more research examining  
588 the cognitive impact of such exposure to conflicting messages (i.e., corrective information followed  
589 by misinformation) on adherence to governmental directives (e.g., around preventive behaviours)  
590 among the public during infectious disease events. Thus far, we know that exposure to conflicting  
591 information around nutrition-related issues has been associated with nutrition confusion, backlash  
592 and decreased performance of healthy behaviours such as fruit and vegetable consumption and  
593 physical activity (Vijaykumar et al., 2021).

594         Further to the above discussions, it is worth highlighting the contribution of gender in the  
595 current study. We did not have a priori predictions surrounding gender or other demographic factors  
596 (e.g., employment) and, thus, did not control for gender distribution in our sampling. Still, inclusion  
597 of gender as a covariate revealed that males performed significantly poorer in our study, i.e., they  
598 were more likely to deem misinformation statements to be more truthful and credible, and self-  
599 reported as being more likely to share the statements with others. Because of the unequal sampling of

600 genders across age groups, we cannot draw heavily on these findings. Still, they do tentatively  
601 indicate that gender contributes to misinformation belief and behaviours, and that young males are at  
602 greater risk of believing in and sharing COVID-19 misinformation. Heightened susceptibility in  
603 young males resonates with existing work that has explicitly investigated the role of gender and other  
604 demographic and socioeconomic factors in COVID-19 misinformation belief (Pickles et al., 2021).  
605 Crucial to our findings, gender could not account for the discussed effects of age and our  
606 intervention 2 manipulation. Building on these tentative findings to explore gender-specific  
607 misinformation effects would be a valuable avenue of future research.

608         What are the consequences of our findings? While not a natural experiment, our study design  
609 was premised on the fact that WhatsApp users can be exposed to the same misinformation once or  
610 multiple times from different sources in their small world network within a short period. In such a  
611 fast-moving informational environment, it would be inappropriate to classify corrective information  
612 as prebunking or debunking, given that it would be virtually impossible to determine who among  
613 millions of users have or have not already been exposed to misinformation. In this context, our  
614 findings show that the benefit of corrective information may be diminished if followed shortly by  
615 misinformation. This is especially pertinent given that misinformation research converges on the  
616 finding that “light-touch” interventions (such as single corrections, infographics, or ‘accuracy  
617 nudges’) are subject to rapid decay over time (Roozenbeek et al., 2021). Thus, our outcomes have  
618 major implications for implementing fact-check/accuracy nudges and other light-touch interventions  
619 in social media environments.

620

## 621 **Conclusion**

622         Reinforcing exposure to corrective information could help maintain the gains from an initial  
623 dose of corrective information. While the exact number of repetitions required to maintain greater  
624 durability of corrective effects has yet to be understood, our findings suggest that a single dose of

625 corrective information is insufficient. Existing work highlights the need for booster doses of  
626 corrective information. Still, our study is one of the few to demonstrate this need a) in the context of  
627 the WHO's official infographics and b) among WhatsApp users. Moreover, our findings cut across  
628 younger and older adults, of whom the latter demonstrated a greater propensity to correctly identify  
629 misinformation and a lower tendency to share it. We suggest that public health agencies like the  
630 WHO leverage ongoing collaborations with the social media industry to ensure that users are  
631 repeatedly exposed to corrective information and gear these interventions among younger adults  
632 whose vulnerability to misinformation is becoming increasingly apparent.

633

#### 634 **Author contributions**

635 Both authors conceptualised the study, designed the methodology, performed data analyses,  
636 and contributed to the writing of the manuscript.

637

#### 638 **Competing interests**

639 The authors have no competing interests to declare. Funding was provided with full  
640 autonomy in research directions and methods.

641 Alhabash, S., & McAlister, A. R. (2015). Redefining virality in less broad strokes: Predicting viral

642 behavioral intentions from motivations and uses of Facebook and Twitter. *New Media &*

643 *Society*, 17(8), 1317–1339. <https://doi.org/10.1177/1461444814523726>

644 Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020a). Evaluating the fake news

645 problem at the scale of the information ecosystem. *Science Advances*, 6(14), eaay3539.

646 Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation

647 of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, 93(1), 59–79.

648 <https://doi.org/10.1177/1077699015606057>

- 649 Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der.  
650 (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking  
651 interventions against COVID-19 misinformation. *Big Data & Society*, 8(1),  
652 20539517211013868. <https://doi.org/10.1177/20539517211013868>
- 653 Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J.,  
654 Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E.  
655 J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., ... Willer, R. (2020). Using  
656 social and behavioural science to support COVID-19 pandemic response. *Nature Human  
657 Behaviour*, 4(5), Article 5. <https://doi.org/10.1038/s41562-020-0884-z>
- 658 Brashier, N. M., & Schacter, D. L. (2020). Aging in an Era of Fake News. *Current Directions in  
659 Psychological Science*, 29(3), 316–323. <https://doi.org/10.1177/0963721420915872>
- 660 Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European  
661 Review of Social Psychology*, 20(1), 49–96. <https://doi.org/10.1080/10463280802643640>
- 662 Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracin, D., Altieri, E., Anoko, J. N., Arcos, R.,  
663 Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S. C., Brooks, I., Bucci, L. M.,  
664 Burzo, S., Czerniak, C., Domenico, M. D., Dunn, A. G., ... Purnat, T. D. (2021). A Public  
665 Health Research Agenda for Managing Infodemics: Methods and Results of the First WHO  
666 Infodemiology Conference. *JMIR Infodemiology*, 1(1), e30979. <https://doi.org/10.2196/30979>
- 667 Carey, J. M., Chi, V., Flynn, D. J., Nyhan, B., & Zeitzoff, T. (2020). The effects of corrective  
668 information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in  
669 Brazil. *Science Advances*, 6(5), eaaw7449. <https://doi.org/10.1126/sciadv.aaw7449>
- 670 Carnahan, D., Bergan, D. E., & Lee, S. (2021). Do Corrective Effects Last? Results from a  
671 Longitudinal Experiment on Beliefs Toward Immigration in the U.S. *Political Behavior*,  
672 43(3), 1227–1246. <https://doi.org/10.1007/s11109-020-09591-9>

- 673 Crime and Security Research Institute. (2020). *Survey of UK Public's Exposure to and Sharing of*  
674 *Coronavirus Disinformation and Fake News*.  
675 [https://www.cardiff.ac.uk/\\_\\_data/assets/pdf\\_file/0006/2399928/OSCAR-Survey-Report-1-](https://www.cardiff.ac.uk/__data/assets/pdf_file/0006/2399928/OSCAR-Survey-Report-1-)  
676 [V2.pdf](https://www.cardiff.ac.uk/__data/assets/pdf_file/0006/2399928/OSCAR-Survey-Report-1-V2.pdf)
- 677 Dror, A. A., Eisenbach, N., Taiber, S., Morozov, N. G., Mizrachi, M., Zigran, A., Srouji, S., & Sela,  
678 E. (2020). Vaccine hesitancy: The next challenge in the fight against COVID-19. *European*  
679 *Journal of Epidemiology*, 35(8), 775–779. <https://doi.org/10.1007/s10654-020-00671-y>
- 680 Dudai, Y., & Eisenberg, M. (2004). Rites of Passage of the Engram: Reconsolidation and the  
681 Lingered Consolidation Hypothesis. *Neuron*, 44(1), 93–100.  
682 <https://doi.org/10.1016/j.neuron.2004.09.003>
- 683 Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P.,  
684 Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief  
685 and its resistance to correction. *Nature Reviews Psychology*, 1(1), Article 1.  
686 <https://doi.org/10.1038/s44159-021-00006-y>
- 687 Edgerly, S. (2017). Seeking Out and Avoiding the News Media: Young Adults' Proposed Strategies  
688 for Obtaining Current Events Information. *Mass Communication and Society*, 20(3), 358–  
689 377. <https://doi.org/10.1080/15205436.2016.1262424>
- 690 Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for  
691 plausible and implausible statements. *Psychonomic Bulletin & Review*, 26(5), 1705–1710.  
692 <https://doi.org/10.3758/s13423-019-01651-4>
- 693 Freelon, D., & Wells, C. (2020). Disinformation as Political Communication. *Political*  
694 *Communication*, 37(2), 145–156. <https://doi.org/10.1080/10584609.2020.1723755>
- 695 Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019a). Fake news on  
696 Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378.  
697 <https://doi.org/10.1126/science.aau2706>

- 698 Guess, A., Nagler, J., & Tucker, J. (2019a). Less than you think: Prevalence and predictors of fake  
699 news dissemination on Facebook. *Science Advances*, 5(1), eaau4586.  
700 <https://doi.org/10.1126/sciadv.aau4586>
- 701 Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect.  
702 *Cognitive Research: Principles and Implications*, 6, 38. [https://doi.org/10.1186/s41235-021-](https://doi.org/10.1186/s41235-021-00301-5)  
703 00301-5
- 704 Herrero-Diz, P., Conde-Jiménez, J., & Reyes de Cózar, S. (2020). Teens' Motivations to Spread  
705 Fake News on WhatsApp. *Social Media + Society*, 6(3), 2056305120942879.  
706 <https://doi.org/10.1177/2056305120942879>
- 707 Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A  
708 subtle reminder triggers integration of new information. *Learning & Memory*, 14(1–2), 47–  
709 53. <https://doi.org/10.1101/lm.365707>
- 710 Jung, E. H., Walsh-Childers, K., & Kim, H.-S. (2016). Factors influencing the perceived credibility  
711 of diet-nutrition information web sites. *Computers in Human Behavior*, 58, 37–47.  
712 <https://doi.org/10.1016/j.chb.2015.11.044>
- 713 Kreps, S. E., & Kriner, D. (2020). *The Covid-19 Infodemic and the Efficacy of Corrections* (SSRN  
714 Scholarly Paper No. 3624517). <https://doi.org/10.2139/ssrn.3624517>
- 715 Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior  
716 experience. *Computers in Human Behavior*, 28(2), 331–339.  
717 <https://doi.org/10.1016/j.chb.2011.10.002>
- 718 Lee, E.-J., & Shin, S. Y. (2021). Mediated Misinformation: Questions Answered, More Questions to  
719 Ask. *American Behavioral Scientist*, 65(2), 259–276.  
720 <https://doi.org/10.1177/0002764219869403>

- 721 Lee, S., Kim, J., & Sung, Y. H. (2022). When infographics work better: The interplay between  
722 temporal frame and message format in e-health communication. *Psychology & Health, 37*(7),  
723 917–931. <https://doi.org/10.1080/08870446.2021.1912342>
- 724 Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through  
725 Inoculation and Prebunking. *European Review of Social Psychology, 32*(2), 348–384.  
726 <https://doi.org/10.1080/10463283.2021.1876983>
- 727 Linden, van der S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S.  
728 (2021). How Can Psychological Science Help Counter the Spread of Fake News? *The*  
729 *Spanish Journal of Psychology, 24*, e25. <https://doi.org/10.1017/SJP.2021.23>
- 730 Lockyer, B., Islam, S., Rahman, A., Dickerson, J., Pickett, K., Sheldon, T., Wright, J., McEachan,  
731 R., Sheard, L., & Group, the B. I. for H. R. C.-19 S. A. (2021). Understanding COVID-19  
732 misinformation and vaccine hesitancy in context: Findings from a qualitative study involving  
733 citizens in Bradford, UK. *Health Expectations, 24*(4), 1158–1167.  
734 <https://doi.org/10.1111/hex.13240>
- 735 Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the  
736 malleability of memory. *Learning & Memory, 12*(4), 361–366.  
737 <https://doi.org/10.1101/lm.94705>
- 738 Maertens, R., Anseel, F., & van der Linden, S. (2020). Combatting climate change misinformation:  
739 Evidence for longevity of inoculation and consensus messaging effects. *Journal of*  
740 *Environmental Psychology, 70*, 101455. <https://doi.org/10.1016/j.jenvp.2020.101455>
- 741 Mutter, S. A., Lindsey, S. E., & Pliske, R. M. (1995). Aging and credibility judgment. *Aging,*  
742 *Neuropsychology, and Cognition, 2*(2), 89–107. <https://doi.org/10.1080/13825589508256590>
- 743 Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and  
744 simulated social media postings: An experimental investigation of source credibility, repeated

- 745 exposure, and presentation format. *Cognitive Research: Principles and Implications*, 5(1), 56.  
746 <https://doi.org/10.1186/s41235-020-00251-4>
- 747 Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative  
748 photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, 19(5), 969–974.  
749 <https://doi.org/10.3758/s13423-012-0292-0>
- 750 Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of*  
751 *General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- 752 Ofcom. (2019). *Children and parents: Media use and attitudes report*.  
753 [https://www.ofcom.org.uk/\\_\\_data/assets/pdf\\_file/0024/134907/Children-and-Parents-Media-](https://www.ofcom.org.uk/__data/assets/pdf_file/0024/134907/Children-and-Parents-Media-Use-and-Attitudes-2018.pdf)  
754 [Use-and-Attitudes-2018.pdf](https://www.ofcom.org.uk/__data/assets/pdf_file/0024/134907/Children-and-Parents-Media-Use-and-Attitudes-2018.pdf)
- 755 Parks, C. M., & Toth, J. P. (2006). Fluency, Familiarity, Aging, and the Illusion of Truth. *Aging,*  
756 *Neuropsychology, and Cognition*, 13(2), 225–253. <https://doi.org/10.1080/138255890968691>
- 757 Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*,  
758 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- 759 Pickles, K., Cvejic, E., Nickel, B., Copp, T., Bonner, C., Leask, J., Ayre, J., Batcup, C., Cornell, S.,  
760 Dakin, T., Dodd, R. H., Isautier, J. M. J., & McCaffery, K. J. (2021). COVID-19  
761 Misinformation Trends in Australia: Prospective Longitudinal National Survey. *Journal of*  
762 *Medical Internet Research*, 23(1), e23805. <https://doi.org/10.2196/23805>
- 763 Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation.  
764 *Behavioural Brain Research*, 84(1), 241–246. [https://doi.org/10.1016/S0166-4328\(96\)00153-](https://doi.org/10.1016/S0166-4328(96)00153-2)  
765 2
- 766 Roozenbeek, J., Freeman, A. L. J., & van der Linden, S. (2021). How Accurate Are Accuracy-Nudge  
767 Interventions? A Preregistered Direct Replication of Pennycook et al. (2020). *Psychological*  
768 *Science*, 32(7), 1169–1178. <https://doi.org/10.1177/09567976211024535>

- 769 Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., Van Der Bles,  
770 A. M., & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19  
771 around the world. *Royal Society Open Science*, 7(10), 201199.
- 772 Song, H., So, J., Shim, M., Kim, J., Kim, E., & Lee, K. (2023). What message features influence the  
773 intention to share misinformation about COVID-19 on social media? The role of efficacy and  
774 novelty. *Computers in Human Behavior*, 138, 107439.  
775 <https://doi.org/10.1016/j.chb.2022.107439>
- 776 Southwell, B. G., Thorson, E. A., & Sheble, L. (2018). *Misinformation and Mass Audiences*.  
777 University of Texas Press.
- 778 Spiers, H. J., & Bendor, D. (2014). Enhance, delete, incept: Manipulating hippocampus-dependent  
779 memories. *Brain Research Bulletin*, 105, 2–7.  
780 <https://doi.org/10.1016/j.brainresbull.2013.12.011>
- 781 Stefanone, M. A., Vollmer, M., & Covert, J. M. (2019). In News We Trust?: Examining Credibility  
782 and Sharing Behaviors of Fake News. *Proceedings of the 10th International Conference on*  
783 *Social Media and Society*, 136–147. <https://doi.org/10.1145/3328529.3328554>
- 784 Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting  
785 inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and*  
786 *Cognition*, 43, 1948–1961. <https://doi.org/10.1037/xlm0000422>
- 787 Umanath, S., & Marsh, E. J. (2014). Understanding How Prior Knowledge Influences Memory in  
788 Older Adults. *Perspectives on Psychological Science*, 9(4), 408–426.  
789 <https://doi.org/10.1177/1745691614535933>
- 790 van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking Formula for Misinformation Treatment in  
791 Public Health Crises: The Effects of Corrective Information Type and Source. *Health*  
792 *Communication*, 35(5), 560–575. <https://doi.org/10.1080/10410236.2019.1573295>

- 793 Vijaykumar, S., Jin, Y., Rogerson, D., Lu, X., Sharma, S., Maughan, A., Fadel, B., de Oliveira Costa,  
794 M. S., Pagliari, C., & Morris, D. (2021). How shades of truth and age affect responses to  
795 COVID-19 (Mis) information: Randomized survey experiment among WhatsApp users in  
796 UK and Brazil. *Humanities and Social Sciences Communications*, 8(1), 1–12.
- 797 Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*,  
798 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- 799 Vraga, E. K., & Bode, L. (2021). Addressing COVID-19 Misinformation on Social Media  
800 Preemptively and Responsively. *Emerging Infectious Diseases*, 27(2), 396–403.  
801 <https://doi.org/10.3201/eid2702.203139>
- 802 Wang, W. Y. (2017). “*Liar, Liar Pants on Fire*”: A New Benchmark Dataset for Fake News  
803 Detection (arXiv:1705.00648). arXiv. <https://doi.org/10.48550/arXiv.1705.00648>
- 804 World Health Organization. (2022). *COVID-19 Mythbusters – World Health Organization*.  
805 [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters)  
806 [busters](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters)
- 807 Zhou, Y., & Shen, L. (2022). Confirmation Bias and the Persistence of Misinformation on Climate  
808 Change. *Communication Research*, 49(4), 500–523.  
809 <https://doi.org/10.1177/00936502211028049>

810

811