The Alan Turing Institute

How do people protect themselves against online misinformation?

Attitudes, experiences and uptake of interventions amongst the UK adult population

Content

Authors

Florence E. Enock[1],

Jonathan Bright[1],

Francesca Stevens[1],

Pica Johansson[1],

Helen Z. Margetts[1,2]

[1] Public Policy Programme, The Alan Turing Institute, The British Library, 96 Euston Road, London. NW1 2DB.
[2] Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford. OX1 3JS.

Corresponding author

Jonathan Bright: jbright@turing.ac.uk

Acknowledgments

Abstract

The prevalence of misinformation online continues to be a global problem that has elicited responses from a variety of stakeholders. Many interventions have been developed and much research has tested their efficacy, often providing promise that effective solutions are possible.

However, far less is known about public support for and uptake of such interventions. In a nationally representative survey of 2000 adults living in the UK, we examine people's engagement with a range of interventions along with misinformation exposure, concerns, susceptibility, and trust in different institutions. While support for and engagement with platform-based solutions and publicly presented initiatives such as downranking, deplatforming, fact check overlays and debunking campaigns is positive, active participation with various important resources is low. Though much work supports their efficacy, just 3% of the population have taken a media literacy course and 7% have used self-help resources such as fact-checking tiplines. Engagement with reporting functions is better, but of the 37% who had flagged misinformation, only 15% were satisfied that appropriate action had been taken.

Our results are important because no matter how effective various interventions are shown to be, they will only be useful if the public is supportive of and engaged with their implementation.

Background

The problem and prevalence of online misinformation

Online misinformation, broadly thought of as content which makes false or misleading claims (Wardle & Derakhshan, 2017), continues to be problematic and widespread. Misinformation has the potential to adversely affect individuals and societies in many ways, including by manipulating public opinion, generating political unrest, influencing individuals' voting behaviours, hindering action on the climate crisis, discouraging people from receiving important vaccines or medications, encouraging people to put themselves at physical risk, and encouraging harm against others based on conspiracies (Banaji & Bhat, 2019; Christie, 2021; Dharawat et al., 2022; Lewandowsky et al., 2017; McPhedran et al., 2023; Meredith & Morse, 2015; Van Bavel et al., 2020; van der Linden, 2022; Vosoughi et al., 2018). As well as creating or reinforcing beliefs that are factually incorrect amongst the population, there are concerns that the prevalence of online misinformation causes general distrust in information more generally, with growing scepticism about the veracity of all information, even that which is factually true and verified (Sanchez & Middlemass, 2022).

The term misinformation may be understood differently across sectors and disciplines, often used interchangeably with terms like disinformation and fake news. Several typologies class disinformation as false information shared with the intent to deceive, with misinformation as false information which may be shared unwittingly and without deceptive intentions (Hernon, 1995; Wardle & Derakhshan, 2017). While use of the term 'fake news' increased at the time of the 2016 presidential election and described false articles designed to mimic those from traditional outlets (Allcott & Gentzkow, 2017; Ecker et al., 2022; Lazer et al., 2018; Pennycook et al., 2018), this term is not used as frequently anymore. For simplicity in this work, we use the term misinformation to cover all categories of false information and do not presuppose an intent to deceive or otherwise.

Online misinformation takes many forms, including fabricated headlines and news stories, audio recordings, images and videos which are misleadingly presented out of context, content which has been manipulated (e.g., using editing software like Photoshop), and, increasingly, audio, image and video content which has been manipulated or created using machine learning methods to alter how a person, object or environment is presented, known as 'deepfakes' (CDEI, 2019; Paris & Donovan, 2019; Pennycook & Rand, 2021; Vraga & Bode, 2020). While misinformation existed long before social media and widespread access to the internet, social media platforms have facilitated its proliferation owing to large, easy-to-reach audiences and environments which can be difficult to monitor at scale (van

der Linden, 2022; Vosoughi et al., 2018). It is difficult to objectively measure how much misinformation exists online because platforms do not always give researchers access to data providing this information, but The World Health Organization declared a global 'infodemic' (an excess of incorrect and deceptive information) in early 2020 (Thomas, 2020; van der Linden, 2022) and a 2019 survey by Ipsos, conducted on behalf of the Centre for International Governance Innovation, found that across 25,000 interviews in over 25 economies, 86% of people online globally believe they've been exposed to false information online. Further work found that concerns are correspondingly high, with 85% of people globally worried about false information online (Ipsos, 2019).

In terms of susceptibility to online misinformation, some work finds that older adults share false news stories on social media more commonly than their younger counterparts and suggests that this may be driven by lower levels of digital literacy (Brashier & Schacter, 2020). There is also work proposing that people are more susceptible to believing misinformation that supports their worldviews and existing belief systems, such as content in line with their political values or content which portrays ingroup members positively (Gawronski et al., 2023; Lewandowsky et al., 2012; Pereira et al., 2023). However, work elsewhere suggests that misinformation susceptibility is more strongly associated with a lack of critical engagement than with biased reasoning (Pennycook & Rand, 2019). Research supporting this approach noted that performance on the Cognitive Reflection Test (an instrument that measures people's aptitude to engage in analytical reasoning) is negatively correlated with the perceived accuracy of fake news, and positively correlated with the aptitude to distinguish false news from real news even for headlines that support individuals' political beliefs (Pennycook et al., 2021; Pennycook & Rand, 2019).

Interventions against misinformation

A wide range of interventions have been (and continue to be) developed in response to the spread of online misinformation, designed across multiple disciplines such as psychology, cognitive science, education and computer science, and implemented by a range of actors including platforms, governments and civil society organisations. A recent review from Johansson and colleagues (2022) synthesises the many current approaches to tackling online misinformation, positioning the available interventions into an analytical framework which considers the stage of the misinformation lifecycle at which they are applied and who is best placed to implement each one, also analysing the interventions for efficacy and feasibility. Under this framework, interventions may aim to **Prepare** people for possible exposure through **Education** (Media literacy courses or Inoculation games) and through **Priming** users (using General warnings or Public awareness campaigns); interventions may aim to **Curb** the spread through **Contextualising** content (using Fact-check labels, Tiplines, Accuracy prompts or Provenance cues), through **Slowing** content (through Demonetisation, Algorithmic downranking or Delisting),

and through **Removal** (by Early stage moderation or Deplatforming); and finally may aim to **Respond** to content by **Correcting** claims (using Debunking techniques or Counterspeech). As such, interventions differ by when they are implemented, how they are implemented, and by which effects of misinformation they hope to tackle (initial exposure, spread, and resulting changes in beliefs and behaviours) (Johansson et al., 2022).

Much research has examined the efficacy, ethical implications and feasibility of implementing the various misinformation interventions (for reviews, see Altay, 2022; Johansson et al., 2022; Roozenbeek et al., 2022). Prior analyses have shown support for interventions aimed at preparing users for exposure through education in the form of media literacy campaigns and inoculation games (Guess et al., 2020; McPhedran et al., 2023; Murrock et al., 2018; Roozenbeek & van der Linden, 2019; Van der Linden et al., 2017), while weaker support for the efficacy of interventions aiming to prime users to be alert to misinformation, such as through general warnings and public awareness campaigns suggests these may work best when folded into wider initiatives (Clayton et al., 2020; Greene & Murphy, 2021; Pantazi et al., 2021; West & Bergstrom, 2021). There is some evidence in support of interventions designed to curb the spread of misinformation such as fact check labels and accuracy prompts (Nieminen & Rapeli, 2019; Nyhan et al., 2020; Pennycook et al., 2021), and interventions which aim to respond to misinformation by correcting false or misleading claims are considered to be important as a last defence within a multi-pronged approach, particularly when they replace a false narrative with an alternative one, and if the correction comes from a trusted source (Chan et al., 2017; Ecker & Antonio, 2021; Jamieson et al., 2021).

While there is a vast body of literature assessing the effectiveness of different interventions on different outcome measures, these studies almost always focus on efficacy rather than on uptake. This means that while we broadly understand what may and may not work, we do not know about people's levels of engagement with and uptake of such interventions outside of controlled research environments such as lab-based experiments and randomised controlled trials with carefully assigned treatment conditions. For example, many studies over the past eight years have examined inoculation games for reducing people's susceptibility to misinformation. In these games, players typically navigate a simulated social media environment while also learning about common methods that purveyors of misinformation rely on. By pre-emptively warning people against misleading tactics and by exposing people to a weakened version, it is claimed that cognitive resistance can be developed against a range of misinformation types and contexts (Basol et al., 2020; Lewandowsky & van der Linden, 2021; McPhedran et al., 2023; Roozenbeek, van der Linden, et al., 2022; Roozenbeek & Linden, 2020; Traberg et al., 2022; Van der Linden et al., 2017). However, while such inoculation approaches show promise in experimental settings, their success will ultimately rely on voluntary participation, and the extent to which people engage or are willing to engage with these is unknown. Similarly, there is

research to show that media literacy courses are effective at increasing people's abilities to detect false news stories, but little is known about the extent to which people actively enrol on such courses in order to learn such skills.

The current research

With so much public discourse taking place online, and at a time of many elections globally, it is crucial to understand not only which interventions against misinformation are effective, but also when and how members of the public are willing to engage with such interventions, and which people are most likely to do so. This is important because no matter how effective they are proven to be, interventions are only useful if people are supportive of their implementation and are willing to engage.

To understand what the UK population are currently doing to protect themselves against the threat of misinformation online, we examine people's awareness of, attitudes towards and engagement with a range of interventions designed to tackle misinformation using a large, up-to-date nationally representative survey of adults in the UK. Using the synthesis of currently available interventions outlined in Johansson et al. (2022), we group eleven interventions into three categories and ask participants a series of questions about each. Our first category is of platform-initiated interventions that are implemented behind the scenes and includes demonetisation, downranking, early moderation and deplatforming. We group these interventions together because these are ones which participants may be aware of but will not have knowingly seen in action. Interventions such as these are difficult to examine in terms of efficacy, but with many platforms implementing such strategies alongside policy support for these approaches (Courchesne et al., 2021; Hiar, 2021; Perez & Hatmaker, 2020; Saltz & Leibowicz, 2021; Telford, 2019), it is important to find out what the public think about their use. We ask members of the UK public whether they are aware of these four interventions and how comfortable they feel with platforms using them.

Our second category is of interventions that are publicly presented to people (but which they need not actively engage in) and includes public awareness campaigns, accuracy prompts, fact check labels and debunking campaigns. We group these interventions together because these are ones which people may have seen without actively seeking them out. We ask participants whether they are aware of each of these four interventions, if they have seen them, and (if they have) whether they followed up on the resources offered.

Our third category is of interventions where active participation is required and includes media literacy courses, inoculation games and self-help resources. We group the interventions this way because these are ones which people must actively seek out and participate in. We ask participants whether they are aware of each of

these three participatory interventions, if they have taken part in them, and whether they would choose to take part in the future. As well as measuring overall proportions of people taking part, we also test which social and demographic characteristics predict engagement.

In addition to measuring attitudes towards and experiences with misinformation interventions, we examine self-reported exposure to online misinformation along with concern about such content, susceptibility to common items of misinformation, and trust in institutions including the UK government, mainstream media and scientists. Further, we examine the extent to which people typically report misinformation online when they see it and which people are most likely to do so.

Methods

Data collection, ethics and open science

Data collection took place online during March 2023 and the survey was created and administered using Qualtrics (https://www.qualtrics.com), with participants recruited through Prolific (https://www.prolific.com). The survey was approved by the Ethics Committee at The Alan Turing Institute, UK (approval number C2105-074). Informed consent was obtained at the start of the survey according to approved ethical procedures. The materials and data will be available open access on publication.

Sample

A total of 1993 participants who completed the survey passed standard checks for data quality and were included in the final sample. The sample was designed to be nationally representative of the population of the United Kingdom across demographic variables of age, gender and ethnicity (using Prolific's representative sample tool). Respondents were aged between 18 and 88 years old, with a mean age of 45.7 (SD = 15.6). A total of 1018 participants identified as female (51%), 963 as male (48%), and 7 as non-binary/third gender (0.4%), with three selecting 'prefer not to say' (0.2%), and two selecting 'prefer to self-describe' (0.1%). The majority were White (1738, 87%), while 150 (7.5%) were Asian or Asian British, 61 (3%) were Black, African, Caribbean, or Black British, and 30 (1.5%) were mixed, multiple or other ethnicities. Ten (<1%) participants selected 'any other ethnic group', whilst four (<1%) chose 'prefer not to say'. Although participants indicated more specific ethnic identities, we have combined them into broader categories to simplify reporting here. 1091 respondents had degree level qualifications (55%), 286 participants had non-degree level qualifications (vocational or similar) (14%), 614 had no qualifications (including completion of secondary school and below) (31%), and 2 participants selected the 'prefer not to say' option (0.1%). As with ethnicity, whilst participants indicated more specific education levels, we combine them into broader categories here.

Survey

Demographics and background questions

For each participant, we collected standard demographic information about age, gender, ethnicity, education and political orientation. Age could be entered as any number with a minimum of 18. For gender, ethnicity and education level, participants were asked to select the option that they felt best described them from a list of standard predefined categories. For political orientation, participants used an unmarked sliding scale (scored from 0-100) to indicate their political ideology from

'extremely left' to 'extremely right' (with 'centre' in the middle). Participants were also asked what kind of device they were using to complete the survey, as well as how many hours of their personal time they typically spend using social media per day. All demographic questions, other than their age, provided participants with a 'prefer not to say' option (this was not included for age as being 18 or over was a requirement to participate).

Trust in government, news organisations and scientists

We asked participants about their level of trust in three different institutions: the UK government, UK-based mainstream news organisations, and UK-based scientists (Do not trust at all/ Do not trust very much/ Trust somewhat/ Trust very much). We also asked participants how much they believe these different groups intend to mislead or misinform the public (Not at all/ Not very much/ Somewhat/ Very much).

Misinformation susceptibility

To measure susceptibility to misinformation for each participant, we presented them with eight statements and asked them to indicate if they thought each one was 'True or more likely to be true', or 'False or more likely to be false'. Four of these items were true (*The benefits of vaccination programmes in the UK vastly outweigh the risks and costs; In some cases, Covid-19 could become a severe disease with respiratory insufficiency requiring intensive care and potentially leading to death; Antarctica is melting because of climate change caused by human activities; Global warming driven by humans is likely to cause an increase in widespread flooding, extreme heat, drought and poverty*), and four were false (*There is a causal link between the MMR (measles, mumps and rubella) vaccine and autism; The 5G network is proven to cause headaches, migraines and dizziness; The coronavirus was created as part of a global effort to enforce mandatory vaccination; The snow and cold weather experienced this winter demonstrates that the world is not at risk from climate change*). These items were combined into a single misinformation recognition score by combining the percentage of correct responses (false items indicated as false, and true items indicated as true) across the eight items for each participant.

Misinformation exposure and concern

To find out how much people are exposed to misinformation online, we asked participants to indicate the extent to which they have witnessed content which they consider to be misinformation in the past on social media platforms (meaning seeing content not directly intended for them), along with the extent to which they have directly received misinformation in the past (meaning seeing content directly intended for them, such as in a direct message). Response options for both scales were: Many times / Occasionally - from time to time / Very rarely – only once or twice / Never / Not sure / Prefer not say. Participants were also asked how concerned they

feel about the spread of misinformation online (Extremely concerned / Fairly concerned / Not very concerned / Not at all concerned / Not sure).

Reporting

To understand the extent to which people report content which they consider to be misinformation, we asked respondents if they have ever disputed or reported ('flagged') an online comment or claim that they had seen which they considered to be misinformation (Yes / No / Not sure). If 'Yes' was selected, respondents were asked whether they were satisfied with the actions taken in response to their report (Yes / No – no action was taken as far as I am aware / No – I am not satisfied with the action taken / Not sure.

Attitudes and experiences with interventions against misinformation

For each of the four 'behind the scenes' interventions, participants were asked if they had heard of the intervention (Yes/ No/ Not Sure) and how comfortable they are with platforms using that type of intervention (Very / Fairly / Not very / Not at all / Not sure).

For each of the 'publicly presented' interventions, participants were asked if they had heard of the intervention, whether they had ever seen the intervention (Yes/ No/ Not Sure), and if they had seen the intervention, whether they had followed up on the resources offered (Yes – every time / Yes – sometimes / No – never / Not sure). If participants indicated they had not followed up on the resources offered, they were asked to provide reasons using a multiple choice list and could choose as many as applied from the following: *Not enough time, Too much effort, I don't think I would learn much from this resource, Difficult to find the resources, Do not trust the information offered, I intended to but forgot, Something else* [free text option], *Not sure* (exclusive).

For each of the 'participatory' interventions, participants were asked if they had heard of the intervention and if they had ever participated in one (Yes/ No/ Not Sure). If participants answered 'No' or 'Not sure' for the latter, they were asked if they thought they would use one in the future (Yes/ No/ Not Sure). If participants said that they would not seek to use the intervention in the future (or were unsure if they would), they were asked to provide reasons using a multiple choice list and could choose as many as applied from the following: *Not enough time, Too much effort, I don't think I would learn much from this resource, Difficult to find the resources, I do not trust the information offered, I might intend to but then forget, Something else* [free text option], *Not sure* (exclusive). If participants said that they had used the intervention before or would seek to do so in the future, they were asked to provide reasons using a multiple choice list and could choose as many as applied from the following: *To be able to better 12ecognize content that may be misinformation, To*

understand how to protect my family and friends from misinformation, To help stop the spread of misinformation in society, The resource has been recommended to me by family and friends, The resource has been recommended to me in an online ad, It was a requirement in my place of work or education (not included for self-help resources where not relevant)*, Something else* [free text option], *Not sure* (exclusive). Descriptions shown to participants for each intervention are included in Table S1 (Supplementary Information).

Procedure

After participants gave their informed consent to take part in the survey, they responded to the background and demographic questions. Following this, participants responded to questions about trust in the UK government, news organisations and scientists, before responding to the eight statements designed to measure misinformation susceptibility. Participants then completed questions about prior exposure to online misinformation, overall concern levels about such content, and reporting behaviours. Next, participants were presented with the questions about attitudes, experiences and uptake of the eleven interventions. At the end of the questions, participants were given an opportunity to provide feedback in a free text box before continuing to the debrief and finally completing the submission and being returned to Prolific for payment. The survey was designed to take approximately 14 minutes to complete and each participant received £2.10 for their time.

Data analysis

We present descriptive statistics for overall proportions of people choosing each response option for exposure to misinformation, concern about misinformation and trust in organisations and we present summaries for misinformation susceptibility scores. We also present descriptive statistics showing proportions of people choosing each response for questions about awareness, comfort with, experience and uptake of each of the eleven interventions of interest. Additionally, we show the top reasons given for engaging or not engaging with each of the publicly presented and participatory interventions.

As well as describing responses to our key survey questions, we use a series of logistic regressions to explain key predictors of engagement with each of the participatory interventions. We test whether age, gender, prior exposure to misinformation, concern about misinformation, social media use per day, misinformation susceptibility and political orientation predicts whether people have used or would use each intervention. Additionally, we test whether these same predictor variables are associated with reporting behaviours. The logistic regressions were run through standard diagnostic tests for multicollinearity and influential values.
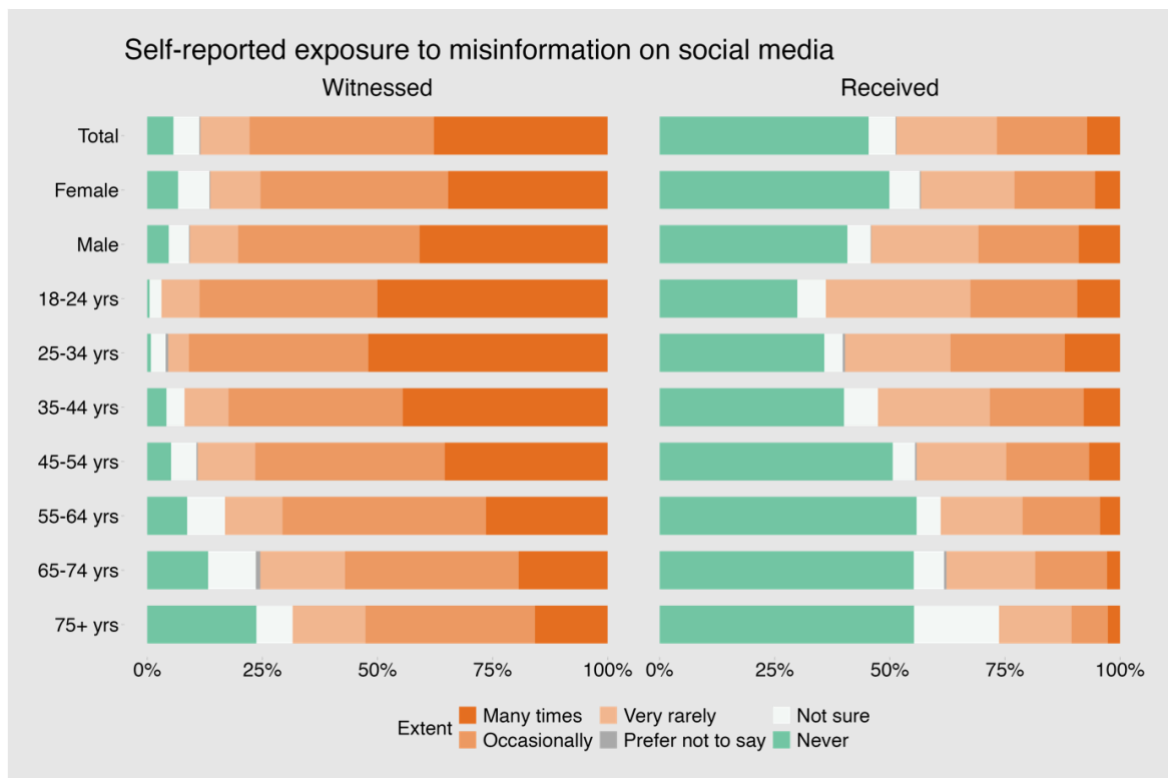
Results

Self-reported exposure to, and concern about, online misinformation

Exposure to misinformation on social media is high: 38% say they have witnessed misinformation on social media many times, **40%** say they have witnessed misinformation on social media occasionally, while **11%** say they have witnessed misinformation on social media very rarely and just **6%** say they have never witnessed misinformation on social media. When respondents who say they never use social media were excluded, only **3%** say they have never witnessed misinformation on social media.

A substantial portion of misinformation is sent directly to individuals on social media: 50% have directly received misinformation on social media at least once. Figure 1 shows self-reported exposure to misinformation broken down by categories for gender and age.

Concern about misinformation is correspondingly high: 86% are fairly or extremely concerned about misinformation on social media. **12%** are not very concerned, and **less than 1%** are not at all concerned.

Figure 1: Self-reported exposure to misinformation on social media broken down by categories for gender and age. The first panel shows the extent to which people have witnessed

misinformation and the second panel shows the extent to which people have directly received misinformation.

## Trust and misinformation susceptibility

People have low trust in UK mainstream news organisations: The majority indicate that they do not trust mainstream news, with **20%** saying they do not trust them at all, and **42%** said they do not trust them very much. **36%** say they trust mainstream news organisations somewhat, but just **3%** indicate they trust them very much. This lack of trust is underpinned by perceived intent - **72%** of people believe the mainstream news intends to mislead them somewhat or very much. Trust in the UK government is also low. A majority said that they do not trust the UK government, with **42%** saying they do not trust the government very much and **32%** saying they do not trust them at all. **25%** trust the government somewhat, but just **2%** say they trust the UK government very much. **81%** say they think UK government intends to mislead somewhat or very much.

However, trust in scientists is higher. **53%** say they trust UK-based scientists somewhat, while **37%** say they trust UK-based scientists very much. Just **10%** say they do not trust scientists very much or at all. Similarly, **89%** say they believe UK scientists do not intend to mislead very much or at all.

Relatively high trust in scientists is reflected in fairly low susceptibility to common items of misinformation relating to scientific themes. For each the four false items, percentages of participants judging them as false ranged between **83%** (link between MMR vaccine and autism) and **91%** (5G is likely to cause ill health effects). For each of the four true items, percentages of participants judging them as true ranged between **87%** (the benefits of vaccination programmes outweigh the risks and costs) and **96%** (in some cases Covid can become a severe disease). While the majority can effectively discern the false from true items, this is not to say susceptibility is not a problem. For example, that 17% believe there is a causal link between the MMR vaccine and autism is likely to have clear harmful consequences for vaccine uptake and population health.

In summary, people report having high exposure to, and high concern about, misinformation on social media, while the majority do not trust the mainstream news. A high proportion of people do trust scientists, and people are generally able to judge the veracity of common misinformation items well. However, a meaningful proportion are susceptible to potentially harmful claims. Figure 2 shows trust and misinformation recognition scores broken down by categories for gender and age.
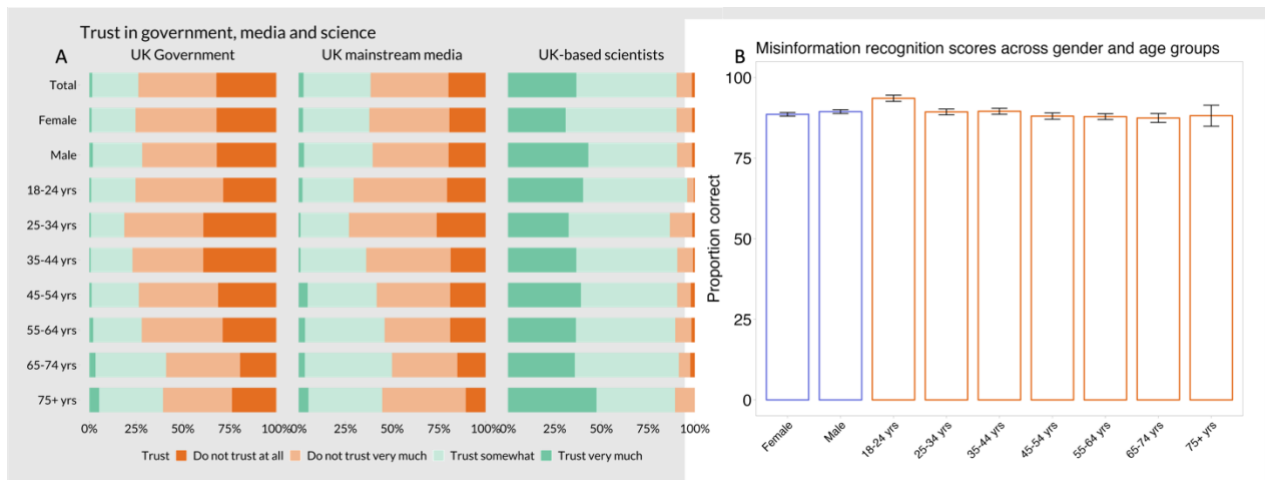
Figure 2: Trust in the UK government, the UK mainstream news, and UK-based scientists broken down by categories for gender and age (A), along with overall misinformation recognition score also broken down by categories for gender and age (B). Misinformation recognition scores are the percentage of correct responses (false items indicated as false, and true items indicated as true) across the eight items.

How do people feel about behind the scenes interventions against misinformation?

Awareness and attitudes

We asked participants about four possible interventions used behind the scenes by platforms to combat the spread of misinformation online: Demonetisation, Downranking, Early moderation, and Deplatforming. We asked people whether they had heard of each intervention, and also how comfortable they feel with their implementation. Approximately half of participants had heard of most of these interventions (Demonetisation – 50%; Downranking – 44%; Early moderation – 45%), with a higher proportion having heard of Deplatforming (66%). Importantly, people are generally comfortable with platforms implementing these kinds of interventions against misinformation, with an average of three quarters of people indicating they are very or fairly comfortable across the four types. **81%** with demonetisation, **74%** with downranking, **76%** with early moderation, and **72%** with deplatforming (Figure 3). **On the whole, people are supportive of platforms intervening behind the scenes, perhaps reflecting the high levels of concern people have about misinformation online.**
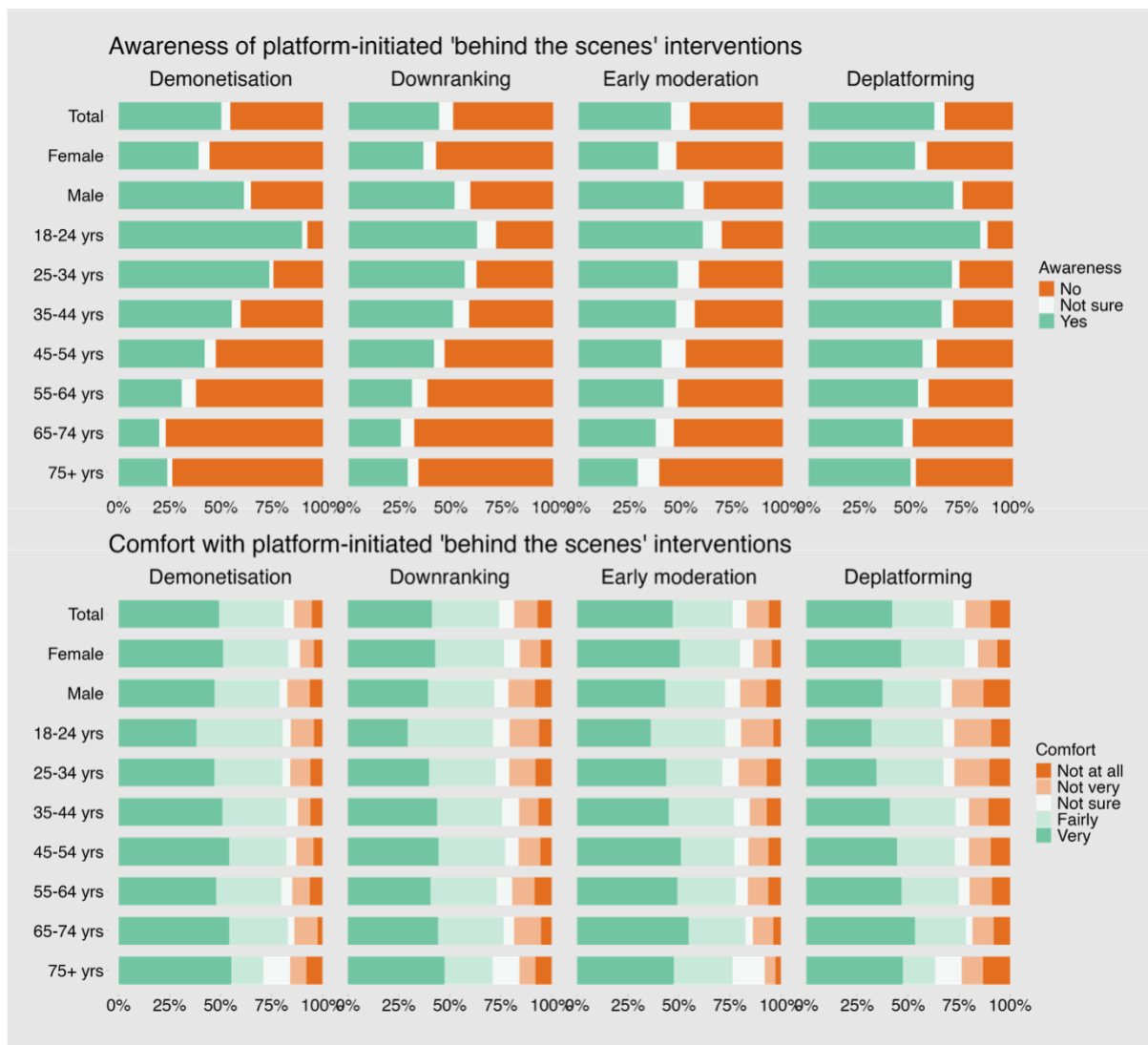
Figure 3: Awareness of and comfort with each platform-initiated 'behind the scenes' intervention broken down by categories for gender and age.

How do people feel about publicly presented interventions?

Awareness, experience and engagement

Approximately half of the participants had heard of most of the publicly presented interventions we asked about (Public awareness campaigns – 56%; Fact checks – 55%; Debunking campaigns – 45%), with a lower proportion having heard of Accuracy prompts (26%), however less than half of the population have seen these interventions in action. People had most commonly seen public awareness campaigns (40%) and fact check labels (45%), while 30% of people have seen debunking campaigns and just 22% have seen online accuracy prompts.

Of the respondents who had seen a public awareness campaign, **32%** said they followed up on the resources offered at least sometimes (response options counted were 'sometimes' and 'every time'). For fact check overlays, this was **65%**; for debunking campaigns, this was **69%**; and for accuracy prompts, this was **75%** (Figure 4). This suggests that at least for fact check overlays, accuracy prompts and debunking campaigns, once people see a publicly presented intervention, they do tend to engage with its message.
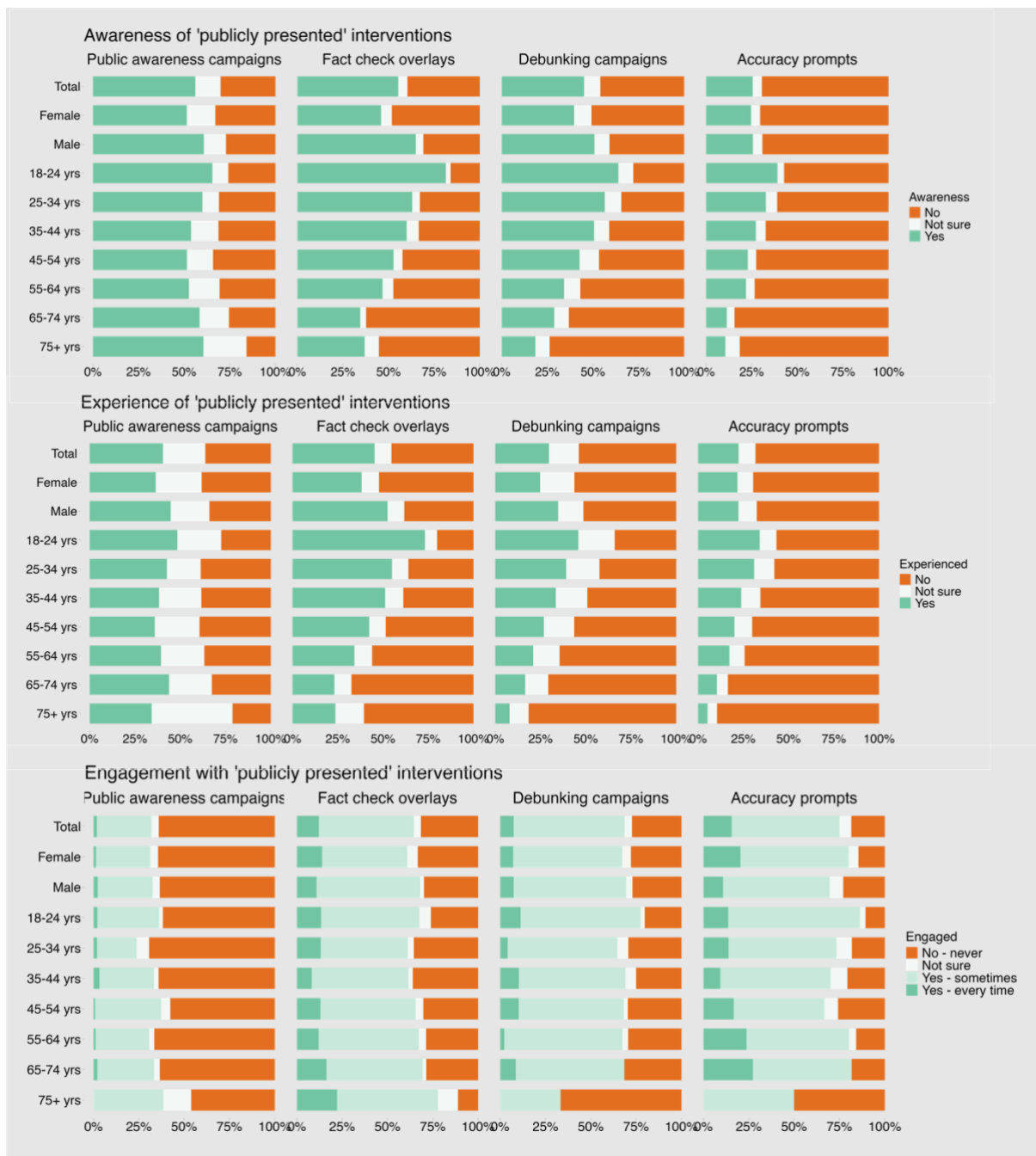
Figure 4: Awareness, experience and engagement with each publicly presented intervention broken down by categories for gender and age. Only people who indicated that they had seen the intervention were asked if they had followed up with the resources offered, meaning percentages for 'engagement' are of this sub-sample rather than the whole survey sample.

Reasons given for not following up on the resources offered

Of the people who said they had seen an intervention but did not follow up on the resources offered, 'I don't think I would learn much from the resource' was the most commonly given reason across all of the publicly presented interventions (48% gave this reason for not following up on public awareness campaigns; 33% for fact check overlays; 34% for debunking campaigns, and 35% for accuracy prompts). Too much time and effort and not trusting the resources were also commonly chosen. Table 1 shows the top four reasons given by participants in response to why they do not follow up on the resources offered in publicly presented interventions.

Table 1: Top four reasons given for not following up on the resources offered for each of the publicly presented interventions. Respondents could choose as many reasons as they felt applied from a multiple-choice list. Only people who indicated that they had seen the intervention but had not followed up were presented with these choices, meaning percentages are of this sub-sample rather than the whole survey sample.

	Public awareness campaigns	Fact check overlays	Debunking campaigns	Accuracy prompts
1	I don't think I would learn much from the resource (48%)	I don't think I would learn much from the resource (33%)	I don't think I would learn much from the resource (34%)	I don't think I would learn much from the resource (35%)
2	Too much time (20%)	Too much effort (26%)	Too much effort (30%)	I do not trust the information offered (31%)
3	Too much effort (19%)	Other (21%)	Too much time (23%)	Too much effort (28%)
4	Other (19%)	I do not trust the information offered (20%)	I do not trust the information offered (21%)	Too much time (15%)

How do people feel about participatory interventions?

Awareness, experience and engagement

Awareness of participatory interventions against misinformation is overall lower than interventions in the other two categories described above - **21%** of people had heard of media literacy courses, **7%** had heard of inoculation games, and **18%** had heard of self-help resources. Just **3%** had taken part in a media literacy course, **5%** had played an inoculation game, and **7%** had used self-help resources before.
Of those who indicated they had *not* taken part in a media literacy course before, **14%** said that they may seek to take part in one in the future, while for inoculation games this was **18%**, and for self-help resources this was **33%** (Figure 5).
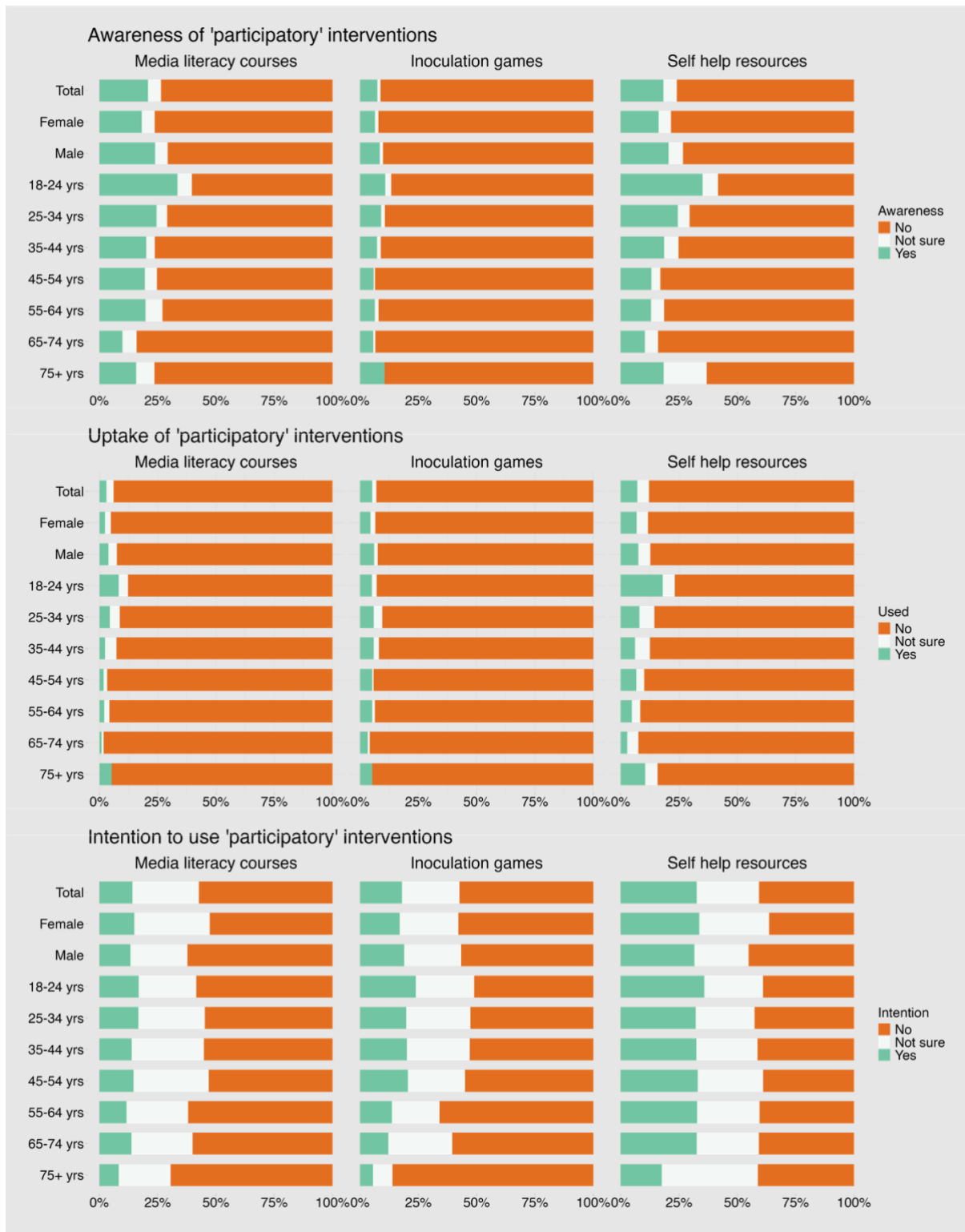
Figure 5: Awareness, uptake and intended uptake of each participatory intervention broken down by categories for gender and age. Only people who indicated that they had not used the intervention before were asked if they would use it in the future, meaning percentages for 'intention to use' are of this sub-sample rather than the whole survey sample.

Reasons given for using each participatory intervention

Of respondents who indicated that they had previously used an intervention, or would use one in the future, 'to be better at recognising content that may be misinformation' was the most commonly chosen reason across the three interventions. 'To help stop the spread of misinformation' and 'To protect family and friends' was also chosen frequently. For inoculation games, a substantial proportion of participants chose 'Other'. The free text responses most commonly centered around having played an inoculation game as part of a Prolific study. Table 2 shows the top four reasons given by participants in response to why they have used or would use each of the participatory interventions.

Table 2: Top four reasons given for using each of the participatory interventions. Respondents could choose as many reasons as they felt applied from a multiple choice list. Only people who indicated that they had used or would use the intervention were presented with these choices, so percentages are of this sub-sample rather than the survey sample as a whole.

	Media literacy courses	Inoculation games	Self-help resources
1	To be able to better recognise content that may be misinformation (83%)	To be able to better recognise content that may be misinformation (70%)	To be able to better recognise content that may be misinformation (75%)
2	To help stop the spread of misinformation in society (64%)	To help stop the spread of misinformation in society (49%)	To help stop the spread of misinformation in society (65%)
3	To understand how to protect my family and friends from misinformation (63%)	To understand how to protect my family and friends from misinformation (46%)	To understand how to protect my family and friends from misinformation (51%)
4	It was a requirement in my place of work or education (17%)	Other (14%)	The resource has been recommended to me by family and friends (8%)

Reasons given for not using each participatory intervention

Of respondents who indicated that they had not and would not use each of these interventions, 'I don't think I would learn much from the resource' was the most commonly given reason across all three. The interventions being too much time and effort were also frequently chosen. For media literacy courses and inoculation games, a substantial proportion of people selected 'Other' and entered free text. For media literacy courses, the free text responses centered around people indicating they do not feel the need to attend because they already feel media literate. For inoculation games, the free text responses centered around people indicating they do not believe the gamified approach would be right for them, and would be more suitable for children. Table 3 shows the top four reasons given by participants in response to why they would not each participatory intervention.

Table 3: Top four reasons given for not using each of the participatory interventions. Respondents could choose as many reasons as they felt applied from a multiple choice list. Only people who indicated that they had not used and would not use the intervention were presented with these choices, percentages are of this sub-sample rather than the survey sample as a whole.

	Media literacy courses	**Inoculation games**	**Self-help resources**
1	I don't think I would learn much from the resource (51%)	I don't think I would learn much from the resource (55%)	I don't think I would learn much from the resource (42%)
2	Too much time (34%)	Too much time (27%)	Too much effort (30%)
3	Too much effort (24%)	Too much effort (25%)	Too much time (23%)
4	Other (14%)	Other (10%)	I do not trust the information offered (12%)

Which people are more likely to actively participate in interventions against misinformation?

To test key predictors of being exposed to each of the participatory interventions, we conducted three logistic regressions (one for each intervention) testing the effects of 10 key predictor variables on whether people said they had ever used each intervention (Yes compared to No/Not sure). We tested the effects of Age (scaled), Gender, Social media use per day (scaled), Political orientation (scaled), Prior exposure to misinformation, Concern about misinformation, Susceptibility to misinformation, Trust in media, Trust in scientists and Trust in government on participation in each intervention. Full regression results can be found in supplementary information (Table S2).

Media Literacy Courses

Social media use per day, prior exposure to misinformation and concern about misinformation were all associated with prior or intended engagement with media literacy courses. Social media use per day was positively associated, with a 32% increase in likelihood of engagement with each increase in one standard deviation of social media use (p<.001). People with prior exposure to misinformation were 80% more likely to engage in such courses (p=.025), and those concerned about misinformation were 155% more likely to engage in such courses (p=.001)[1].

Inoculation games

Social media use per day, prior exposure to misinformation and concern about misinformation were all associated with prior or intended engagement with inoculation games. Social media use per day was positively associated, with a 22% increase in likelihood of engagement with each increase in one standard deviation of social media use (p=.001). People with prior exposure to misinformation were 113% more likely to engage with these games (p=.002), and those concerned about misinformation were 105% more likely to engage (p=.003)[2].

Self-help resources

Social media use per day, prior exposure to misinformation, concern about misinformation, trust in the news and trust in scientists were all associated with prior or intended engagement with self-help resources. Social media use per day was positively associated, with a 14% increase in likelihood of engagement with each increase in one standard deviation of social media use (p=.02). People with prior exposure to misinformation were 156% more likely to use such resources (p<.001), and those concerned about misinformation were 151% more likely to use these resources (p<.001). Those who trust UK-based mainstream news organisations were 39% more likely to use self-help resources than those who do not trust news organisations (p=.003), and those who trust scientists were 73% more likely to use self-help resources than those who do not trust scientists (p=.007)[3].

In summary, across all participatory interventions, social media use per day, prior exposure to misinformation and concern about misinformation were all positively associated with prior or intended participation. Trust in the news and trust in scientists were additional predictors of engagement with self-help resources, with those higher in trust more likely to use these resources.

To what extent do people typically report misinformation when they think they see it?

37% of respondents had previously reported content which they considered to be misinformation online. When only people that had reported witnessing misinformation online before were in the analysis, **42%** had reported such content, while the majority (53%) had not.

Of the people who had reported content which they considered to be misinformation, only **15%** were satisfied with the actions taken in response to the report. 62% said no action was taken as far as they were aware, and 11% said they were not satisfied with the action taken (a further 11% were unsure).

Who is more likely to engage with reporting perceived misinformation online?

We used a logistic regression to test key predictors of reporting behaviours (Yes/No & Not sure) using the same 10 predictors as listed for the participatory intervention engagement section. Full regression results can be found in supplementary information (Table S3).

Age, Social media use per day, Political orientation, Prior exposure to misinformation, Concern about misinformation, and Trust in news were all associated with reporting misinformation. Age showed a negative association with younger people more likely to report – there was a 19% increase in likelihood of reporting misinformation with every decrease in one standard deviation of age (p<.001). Social media use per day was positively associated with people using social media more more likely to report – there was a 51% increase in likelihood of reporting for each increase in one standard deviation of social media use per day (p<.001). The more left wing participants were, the greater the likelihood of reporting, with an 18% increase in likelihood with each increase of one standard deviation to the left of the political orientation scale (p<.001) and those concerned about misinformation were 202% more likely to report it (p<.001). Finally, those indicating trust in news organisations were 76% more likely to report than those not trusting the news (p=.018)[4].

---

[4] When 67 influential values were removed, Exposure to misinformation became insignificant.

Discussion

The prevalence of misinformation online continues to be a global problem that has elicited responses from researchers, policymakers, civil society and industry. Many interventions have been developed by a variety of stakeholders to tackle online misinformation, which vary across factors such as which effects of misinformation they hope to target, when in the misinformation lifecycle they are implemented, and who they are implemented by. Much research over the last decade has focused on assessing the efficacy of a range of misinformation interventions, often providing promise that effective solutions do exist, particularly when combined to create a multi-pronged approach (Altay, 2022; Johansson et al., 2022; Roozenbeek, Suiter, et al., 2022). However, far less is known about public attitudes towards and uptake of these interventions outside of controlled research environments. No matter how effective they are thought to be, interventions designed to tackle misinformation will only be useful if the public is supportive of their implementation. As such, we conducted a nationally representative survey of 2000 adults living in the UK to find out what people are currently doing to protect themselves against the threat of misinformation online. We examined people's awareness, attitudes and engagement with a range of interventions, along with general susceptibility to misinformation and trust in different institutions.

Broadly, we found that self-reported exposure to online misinformation is high, with 94% of the UK population having witnessed content that they perceive to be misinformation on social media at least once (97% when only social media users are analysed) and half of the population having been directly targeted by such content, supporting findings elsewhere suggesting exposure to such content globally is high (Ipsos, 2023; Ofcom, 2022; Thomas, 2020). While a 2019 poll by Ipsos found that 86% of people online worldwide believe they've been exposed to false information online, our even higher proportion may be explained by the inclusion of only a UK sample (where exposure may be higher or where people may be more aware of their exposure), or may point to a rise in the problem over the past five years. We also found that a large proportion of the UK public are concerned about the spread of misinformation, with 86% saying they are fairly or extremely concerned, a result mirroring global proportions in recent work elsewhere (Ipsos, 2023).

Concerns about misinformation are reflected in low trust in the mainstream media and in the government, with 62% saying they do not trust mainstream news organisations very much or at all, and 72% indicating they believe these organisations intend to mislead them. This may suggest that the known prevalence of misinformation may produce general distrust in our information environments overall, which demonstrates the ever-increasing need to equip people with the tools to understand how to analyse and verify content they are presented with. More optimistically, trust in scientists is higher, with 90% saying they trust UK based scientists somewhat or very much. This is promising for the rollout of interventions which rely on science as an authority, such as public awareness campaigns reminding people about vaccination. Despite high exposure to misinformation, this trust in science is reflected in relatively low susceptibility to common items of climate and health misinformation, with a majority able to tell true from false information. This is not to say exposure is without problems – we find that 17% of people in the UK believe there is a causal link between the MMR vaccine and autism, a number mirrored elsewhere (Fombonne et al., 2020; YouGov, 2021) and high enough to have harmful consequences for vaccine uptake and population health.

High overall concern in the population about the spread of misinformation is reflected in support for platform-initiated interventions. While about half the population on average had heard of the four 'behind the scenes' interventions that we asked about, the majority were comfortable with their implementation (72% or above for all four). This supports work elsewhere finding that people are overwhelmingly in support of action from both government and from social media platforms to tackle online harms (Enock et al., 2023; Ipsos, 2023). With many platforms implementing such strategies alongside policy support for these approaches (Courchesne et al., 2021; Hiar, 2021; Perez & Hatmaker, 2020; Saltz & Leibowicz, 2021; Telford, 2019), it is reassuring to find their implementation is supported by the public.

For most of the 'publicly presented' interventions (fact check overlays, accuracy prompts and debunking campaigns), engagement is promising when people see them. There is empirical support for the effectiveness of these interventions at reducing people's susceptibility to believing false information, reducing sharing intentions, and correcting incorrect beliefs formed as a result of exposure (Chan et al., 2017; Johansson et al., 2022; Nieminen & Rapeli, 2019; Pennycook & Rand, 2022) and we add support for their rollout by showing promising levels of public engagement. However, only about half the population have seen public awareness campaigns, fact check overlays and debunking campaigns, and about a quarter have seen accuracy prompts, far fewer than have seen misinformation. Increasing the salience and reach of these types of interventions is an important step in fully utilising their potential. We note here that people were less likely to engage with the resources offered by public awareness campaigns than the other publicly presented interventions which may be because these can be seen in offline as well as online contexts when people are not immediately available to follow up. Further work should investigate in more detail possible reasons here and how to enhance engagement.

Reasons given for not engaging with participatory interventions centered on people being unconvinced of their utility and perceiving them as taking too much time and effort. Emphasising the usefulness of these participatory interventions and making them accessible and sensitive to people's time constraints may therefore enhance uptake. We found that people who use social media more, report having seen misinformation in the past and who are concerned about misinformation are more likely to actively participate in interventions designed to tackle it, suggesting uptake should be particularly encouraged amongst those who are less digitally engaged. Rather than solely focusing on researching efficacy (summarised in Johansson et al., 2022), it is crucial for further work to focus on designing effective ways to engage people with these participatory interventions to ensure members of the public are equipped with the right tools needed to protect themselves against misinformation and its negative effects.

User reporting of misinformation online (known as 'flagging') is considered to be important for monitoring and tackling the problem at scale (Allen et al., 2021; Coleman, 2021; Crawford & Gillespie, 2016). Our results show that though a substantial proportion of people who have seen misinformation have used reporting mechanisms to flag it, this is still a minority (42%). More concerningly, when people have reported misinformation, an overwhelming majority have been unsatisfied with actions taken in response to their report, with 62% saying no action had been taken. Dissatisfaction with reporting mechanisms for online harms has been recently reported elsewhere (Bright et al., 2024) and must be addressed because it is likely that people will only use such safety features if they believe them to be effective. Platforms must work to make safety tools such as reporting more salient, accessible and effective for them to be a useful tool in combatting the spread of misinformation (Bright et al., 2024; Molina & Sundar, 2022; Naab et al., 2018). These results highlight the importance of new requirements set out in the Online Safety Act for larger platforms to ensure that all users have straightforward access to effective safety technology (Ofcom, 2023).

Our survey report provides novel, up-to-date evidence surrounding the UK public's current exposure to online misinformation, associated concerns, how people feel about and engage with a range of interventions designed to tackle the problem, along with their trust in key institutions. The work benefits from the inclusion of a large, nationally representative sample, allowing us to make generalisations across the population, and contributes to the existing literature on efficacy of interventions by providing important new insights about public sentiments and engagement with currently available initiatives. However, it is important to acknowledge limitations in this work and acknowledge outstanding research questions that were not possible to answer. Our findings are based solely on self-report data, meaning that conclusions drawn about, for example, the extent of exposure, are based on whether people think they have seen misinformation, rather than objectively quantifying how much misinformation exists online. For a full picture, these results could be interpreted alongside other data sources such as large-scale analyses of content on social media (e.g., Del Vicario et al., 2016). Additionally, our sample included only adults, and future work will benefit from understanding exposure to misinformation and engagement with interventions designed to protect against it in children and adolescents, who may be particularly at risk of suffering the adverse effects (Howard et al., 2021).

Conclusion

Overall, our findings reiterate that exposure to misinformation amongst the population is very high and that most people are concerned about this problem, shown in general support for platform-initiated interventions such as algorithmic downranking and deplatforming. While the public are generally likely to engage with information offered in publicly presented interventions like fact check overlays and debunking campaigns when they see them, that most people have not seen initiatives like these before suggests that more should be done to increase their salience and reach. Most importantly, low awareness and uptake of participatory interventions like media literacy courses and self-help resources shows that it is crucial for stakeholders to engage people with their use, particularly given the body of work showing these initiatives are effective. Similarly, it is important for platforms to provide users with effective reporting mechanisms to increase people's likelihood of flagging misinformation when they see it. With the increasing availability of technologies that can quickly and convincingly create and spread false content online, it is critical that the public are equipped with the right tools to protect themselves against the spread of misinformation and its adverse effects.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. **https://doi.org/10.1257/jep.31.2.211**

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances*, *7*(36), eabf4393.

Altay, S. (2022). *How Effective Are Interventions Against Misinformation?* **https://doi.org/10.31234/osf.io/sm3vk**

Banaji, S., & Bhat, R. (2019, November 11). WhatsApp Vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. *Media@LSE.* **https://blogs.lse.ac.uk/medialse/2019/11/11/whatsapp-vigilantes-an-exploration-of-citizen-reception-and-circulation-of-whatsapp-misinformation-linked-to-mob-violence-in-india/**

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News. *Journal of Cognition*, *3*(1), 2. **https://doi.org/10.5334/joc.91**

Brashier, N. M., & Schacter, D. L. (2020). Aging in an Era of Fake News. *Current Directions in Psychological Science*, *29*(3), 316–323. **https://doi.org/10.1177/0963721420915872**

Bright, J., Enock, F. E., Johansson, P., Margetts, H. Z., & Stevens, F. (2024). *Understanding engagement with platform safety technology for reducing exposure to online harms* (arXiv:2401.01796). arXiv. **http://arxiv.org/abs/2401.01796**

Bulger, M., & Davison, P. (2018). The Promises, Challenges, and Futures of Media Literacy. *Journal of Media Literacy Education*, *10*(1), 1–21. **https://doi.org/10.23860/JMLE-2018-10-1-1**

Case Study: Learn to Discern in Jordan. (n.d.). Retrieved 21 December 2022, from **https://www.irex.org/project/learn-discern-l2d-media-literacy-training**

CDEI. (2019). *Deepfakes and Audiovisual Disinformation*. Centre for Data Ethics and Innovation. **https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation**

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation.

Psychological Science, *28*(11), 1531–1546.
https://doi.org/10.1177/0956797617714579

Chernyavskiy, A., Ilvovsky, D., & Nakov, P. (2021). WhatTheWikiFact: Fact-Checking Claims Against Wikipedia. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 4690–4695. **https://doi.org/10.1145/3459637.3481987**

Christie, L. (2021). *COVID-19 vaccine misinformation*. **https://post.parliament.uk/covid-19-vaccine-misinformation/**

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., & Morgan, E. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095.

Coleman, K. (2021, January 25). *Introducing Birdwatch, a community-based approach to misinformation*. **https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation**

Courchesne, L., Ilhardt, J., & Shapiro, J. N. (2021). Review of social science research on the impact of countermeasures against influence operations. *Harvard Kennedy School Misinformation Review*. **https://doi.org/10.37016/mr-2020-79**

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, *18*(3), 410–428. **https://doi.org/10.1177/1461444814543163**

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. **https://doi.org/10.1073/pnas.1517441113**

Dharawat, A., Lourentzou, I., Morales, A., & Zhai, C. (2022). Drink Bleach or Do What Now? COVID-HeRA: A Study of Risk-Informed Health Decision Making in the Presence of COVID-19 Misinformation. *Proceedings of the International AAAI Conference on Web and Social Media*, *16*, 1218–1227. **https://doi.org/10.1609/icwsm.v16i1.19372**

Ecker, U. K. H., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, *49*(4), 631–644. **https://doi.org/10.3758/s13421-020-01129-y**

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of

misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), Article 1. **https://doi.org/10.1038/s44159-021-00006-y**

Enock, F., Johansson, P., Bright, J., & Margetts, H. Z. (2023). Tracking experiences of online harms and attitudes towards online safety interventions: Findings from a large-scale, nationally representative survey of the british public. *Nationally Representative Survey of the British Public (March 21, 2023)*. **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4416355**

Fombonne, E., Goin-Kochel, R. P., O'Roak, B. J., Abbeduto, L., Aberbach, G., Acampado, J., Ace, A. J., Albright, C., Alessandri, M., Amaral, D. G., Amatya, A., Anglo, C., Annett, R. D., Arriaga, I., Ashley, R., Astrovskaya, I., Baalman, K., Baer, M., Bahl, E., … Zick, A. (2020). Beliefs in vaccine as causes of autism among SPARK cohort caregivers. *Vaccine*, *38*(7), 1794–1803. **https://doi.org/10.1016/j.vaccine.2019.12.026**

Gabriel, Sanchez, & Middlemass, K. (2022). *Misinformation is eroding the public's confidence in democracy* [Commentary]. Brookings. **https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in-democracy/**

Gawronski, B., Ng, N. L., & Luke, D. M. (2023). Truth sensitivity and partisan bias in responses to misinformation. *Journal of Experimental Psychology: General*, *152*(8), 2205–2236. **https://doi.org/10.1037/xge0001381**

Greene, C. M., & Murphy, G. (2021). Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. *Journal of Experimental Psychology: Applied*, *27*(4), 773–784. **https://doi.org/10.1037/xap0000371**

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545.

Hernon, P. (1995). Disinformation and misinformation through the internet: Findings of an exploratory study. *Government Information Quarterly*, *12*(2), 133–139. **https://doi.org/10.1016/0740-624X(95)90052-7**

Hiar, C. (2021, October 8). Google Bans Ads That Spread Climate Misinformation. *Scientific American*. **https://www.scientificamerican.com/article/google-bans-ads-that-spread-climate-misinformation/**

Howard, P. N., Neudert, L.-M., Prakash, N., & Vosloo, S. (2021). Digital misinformation/disinformation and children. *UNICEF. Retrieved on February*, *20*, 2021.

Ipsos. (2019). *Fake News: A Global Epidemic Vast Majority (86%) of Online Global Citizens Have Been Exposed to it.* Ipsos.

Ipsos. (2023). *A global survey on the impact of online disinformation and hate speech.* UNESCO. **https://www.unesco.org/sites/default/files/medias/fichiers/2023/11/unesco_ipsos_survey.pdf**

Jamieson, K. H., Romer, D., Jamieson, P. E., Winneg, K. M., & Pasek, J. (2021). The role of non–COVID-specific and COVID-specific factors in predicting a shift in willingness to vaccinate: A panel study. *Proceedings of the National Academy of Sciences, 118*(52), e2112266118. **https://doi.org/10.1073/pnas.2112266118**

Johansson, P., Enock, F., Hale, S., Vidgen, B., Bereskin, C., Margetts, H., & Bright, J. (2022). *How can we combat online misinformation? A systematic overview of current interventions and their efficacy* (arXiv:2212.11864). arXiv. **http://arxiv.org/abs/2212.11864**

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. **https://doi.org/10.1126/science.aao2998**

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. **https://doi.org/10.1016/j.jarmac.2017.07.008**

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). *Misinformation and Its Correction.* **https://doi.org/10.1177/1529100612451018**

Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology, 32*(2), 348–384. **https://doi.org/10.1080/10463283.2021.1876983**

McPhedran, R., Ratajczak, M., Mawby, M., King, E., Yang, Y., & Gold, N. (2023). Psychological inoculation protects against the social media infodemic. *Scientific Reports, 13*(1), Article 1. **https://doi.org/10.1038/s41598-023-32962-1**

Meredith, M., & Morse, M. (2015). *The Politics of the Restoration of Ex-Felon Voting Rights: The Case of Iowa* (SSRN Scholarly Paper 3872752). **https://papers.ssrn.com/abstract=3872752**

Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: Effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, *27*(4), zmac010.

Murrock, E., Amulya, J., Druckman, M., & Liubyva, T. (2018). Winning the war on state-sponsored propaganda: Results from an impact study of a Ukrainian news media and information literacy program. *Journal of Media Literacy Education*, *10*(2), 53–85.

Naab, T. K., Kalch, A., & Meitz, T. G. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society*, *20*(2), 777–795.

Nieminen, S., & Rapeli, L. (2019). Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-checking Literature. *Political Studies Review*, *17*(3), 296–309. **https://doi.org/10.1177/1478929918786852**

Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*, *42*(3), 939–960. **https://doi.org/10.1007/s11109-019-09528-x**

Ofcom. (2022). *Online Nation 2022 Report*.

Ofcom. (2023). *Ofcom's approach to implementing the Online Safety Act*. Ofcom. **https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation**

Pantazi, M., Hale, S., & Klein, O. (2021). Social and Cognitive Aspects of the Vulnerability to Political Misinformation. *Political Psychology*, *42*(S1), 267–304. **https://doi.org/10.1111/pops.12797**

Pappas, S. (2022). Fighting fake news in the classroom. *Https://Www.Apa.Org*. **https://www.apa.org/monitor/2022/01/career-fake-news**

Paris, B., & Donovan, J. (2019). *Deepfakes and Cheap Fakes*. Data & Society. **https://datasociety.net/library/deepfakes-and-cheap-fakes/**

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology. General*, *147*(12), 1865–1880. **https://doi.org/10.1037/xge0000465**

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), Article 7855. **https://doi.org/10.1038/s41586-021-03344-2**

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. **https://doi.org/10.1016/j.cognition.2018.06.011**

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. **https://doi.org/10.1016/j.tics.2021.02.007**

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), Article 1. **https://doi.org/10.1038/s41467-022-30073-5**

Pereira, A., Harris, E., & Van Bavel, J. J. (2023). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, *26*(1), 24–47. **https://doi.org/10.1177/13684302211030004**

Perez, S., & Hatmaker, T. (2020, November 5). Facebook blocks hashtags for #sharpiegate, #stopthesteal election conspiracies. *TechCrunch*. **https://techcrunch.com/2020/11/05/facebook-blocks-sharpiegate-hashtag-election-conspiracies/**

Roozenbeek, J., & Linden, S. van der. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. *Harvard Kennedy School Misinformation Review*. **https://doi.org/10.37016/mr-2020-47**

Roozenbeek, J., Suiter, J., & Culloty, E. (2022). *Countering Misinformation: Evidence, Knowledge Gaps, and Implications of Current Interventions* [Preprint]. PsyArXiv. **https://doi.org/10.31234/osf.io/b52um**

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), Article 1. **https://doi.org/10.1057/s41599-019-0279-9**

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), eabo6254. **https://doi.org/10.1126/sciadv.abo6254**

Saltz, E., & Leibowicz, C. (2021). *Fact-Checks, Info Hubs, and Shadow-Bans: A Landscape Review of Misinformation Interventions*. Partnership on AI. **https://partnershiponai.org/intervention-inventory/**

Telford, T. (2019, February 22). Pinterest is blocking search results about vaccines to protect users from misinformation. *Washington Post*. **https://www.washingtonpost.com/business/2019/02/21/pinterest-is-blocking-**

all-vaccine-related-searches-all-or-nothing-approach-policing-health-misinformation/

Thomas, Z. (2020, February 13). WHO says fake coronavirus claims causing 'infodemic'. *BBC News.* **https://www.bbc.com/news/technology-51497800**

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 136–151.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), Article 5. **https://doi.org/10.1038/s41562-020-0884-z**

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), Article 3. **https://doi.org/10.1038/s41591-022-01713-6**

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. **https://doi.org/10.1126/science.aap9559**

Vraga, E. K., & Bode, L. (2020). Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication*, *37*(1), 136–144. **https://doi.org/10.1080/10584609.2020.1716500**

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking.* Council of Europe.

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences*, *118*(15), e1912444117. **https://doi.org/10.1073/pnas.1912444117**

YouGov. (2021). *The Economist/YouGov Poll.* YouGov. **https://docs.cdn.yougov.com/w2zmwpzsq0/econTabReport.pdf**

Supplementary Information

Table S1: Full list of interventions included in the survey grouped into the three main categories, along with descriptions for each that were given to participants.

	Intervention	**Description given**
Behind the scenes	Demonetisation	*Demonetising content means that publishers of misinformation can no longer make money from it, for example through adverts.*
	Downranking	*Downranking content using algorithms can make the content appear less frequently on people's newsfeeds or be shown to fewer users.*
	Early moderation	*Early moderation involves preventing certain types of content from being uploaded.*
	Deplatforming	*Deplatforming removes a user or group from a platform when they post content classed as misinformation.*
Publicly presented	Public awareness campaigns	*Aim to raise public awareness about the prevalence of misinformation and the harm that believing and sharing such content may cause. Sometimes these campaigns encourage people to question what they read online. For example, in 2020, the World Health Organisation (WHO) launched 'Stop the Spread' to try and counter misinformation about Covid-19.*
	Accuracy prompts	*Are sometimes used by social media platforms to encourage people to pause before liking or sharing content to consider its truthfulness. For example, on some platforms, users are sometimes asked if they would like to read a full article before sharing a headline on social media.*
	Fact check labels	*Partially or fully overlay content and usually warn users that claims made in the content have been disputed by third-party fact-checkers, sometimes offering links to more information.*

		For example, across some pieces of content on social media you may see written: 'This claim has been disputed by third-party fact-checkers'.
	Debunking campaigns	*Aim to correct false beliefs that are brought about as a result of exposure to misinformation. They work by countering claims made in misinformation with detailed factual explanations. For example, some influencers take to YouTube or other platforms to provide explainers about health-related misinformation.*
Participatory	Media literacy courses	*Aim to equip people with skills to help them critically evaluate content, recognise content that may be misinformation, and reduce susceptibility to believing and sharing such content. These courses may sometimes be offered through places of education or work.*
	Inoculation games	*Are short games that people of all ages can play online to try and help them learn common signs of misinformation. For example, in the 'Bad News' game, players must take on the role of a fake news tycoon and try and gain as many followers as they can.*
	Self-help resources	*Aim to allow individuals to investigate the truthfulness of a claim or gather additional context relating to something they have seen online, for example through social media tiplines, chatbots and fact-check databases. For example, WhatsApp and some other platforms have started misinformation 'tiplines' where users can forward content that they suspect contains misinformation for further resources about the content.*

Table S2: Full regression outputs for predictors of engagement with each participatory intervention.

	Predictor variable	Estimate	SE	p
Media Literacy Courses	Age (scaled)	-0.078	0.069	0.260
	Gender (binary)	0.042	0.124	0.733
	Social media per day (scaled)	0.278	0.065	0.000
	Political orientation (scaled)	-0.087	0.070	0.214
	Prior exposure (binary)	0.588	0.262	0.025
	Concern (binary)	0.937	0.241	0.000
	Susceptibility (binary)	0.267	0.165	0.105
	Trust in news (binary)	0.024	0.142	0.866
	Trust in government (binary)	0.293	0.159	0.065
	Trust in scientists (binary)	0.152	0.240	0.527
Inoculation games	Age (scaled)	-0.079	0.062	0.201
	Gender (binary)	0.217	0.111	0.051
	Social media per day (scaled)	0.196	0.060	0.001
	Political orientation (scaled)	-0.083	0.063	0.187
	Prior exposure (binary)	0.755	0.239	0.002
	Concern (binary)	0.717	0.199	0.000
	Susceptibility (binary)	0.018	0.154	0.905

	Predictor variable			
	Trust in news (binary)	-0.085	0.127	0.503
	Trust in government (binary)	0.062	0.145	0.669
	Trust in scientists (binary)	0.419	0.228	0.066
Self-help resources	Age (scaled)	-0.045	0.054	0.405
	Gender (binary)	-0.054	0.097	0.577
	Social media per day (scaled)	0.127	0.054	0.019
	Political orientation (scaled)	0.041	0.055	0.450
	Prior exposure (binary)	0.940	0.187	0.000
	Concern (binary)	0.918	0.166	0.000
	Susceptibility (binary)	-0.261	0.136	0.055
	Trust in news (binary)	0.327	0.109	0.003
	Trust in government (binary)	0.132	0.124	0.287
	Trust in scientists (binary)	0.547	0.203	0.007

Table S3: Full regression outputs for predictors of engagement with reporting misinformation

	Predictor variable	Estimate	SE	p
Reporting	Age (scaled)	-0.209	0.057	0.000
	Gender (binary)	0.140	0.103	0.173
	Social media per day (scaled)	0.415	0.058	0.000

Political orientation (scaled)	-0.199	0.058	0.001
Prior exposure (binary)	2.345	0.350	0.000
Concern (binary)	1.106	0.182	0.000
Susceptibility (binary)	-0.054	0.142	0.703
Trust in news (binary)	-0.279	0.118	0.018
Trust in government (binary)	-0.143	0.137	0.293
Trust in scientists (binary)	0.049	0.195	0.802