nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01641-6

Understanding and combatting misinformation across 16 countries on six continents

Received: 6 April 2022

Accepted: 24 May 2023

Published online: 29 June 2023

Check for updates

Antonio A. Arechar[1,2,3], Jennifer Allen[2], Adam J. Berinsky[4], Rocky Cole[5], Ziv Epstein[2,6], Kiran Garimella[7], Andrew Gully[5], Jackson G. Lu[2], Robert M. Ross[8], Michael N. Stagnaro[2], Yunhao Zhang[2], Gordon Pennycook[9,10] & David G. Rand[2,11,12]

The spread of misinformation online is a global problem that requires global solutions. To that end, we conducted an experiment in 16 countries across 6 continents (N = 34,286; 676,605 observations) to investigate predictors of susceptibility to misinformation about COVID-19, and interventions to combat the spread of this misinformation. In every country, participants with a more analytic cognitive style and stronger accuracy-related motivations were better at discerning truth from falsehood; valuing democracy was also associated with greater truth discernment, whereas endorsement of individual responsibility over government support was negatively associated with truth discernment in most countries. Subtly prompting people to think about accuracy had a generally positive effect on the veracity of news that people were willing to share across countries, as did minimal digital literacy tips. Finally, aggregating the ratings of our non-expert participants was able to differentiate true from false headlines with high accuracy in all countries via the 'wisdom of crowds'. The consistent patterns we observe suggest that the psychological factors underlying the misinformation challenge are similar across different regional settings, and that similar solutions may be broadly effective.

The spread of misinformation online has become a major cause of concern in recent years. Although the 2016 United States Presidential Election and British 'Brexit' referendum triggered an explosion of academic research on 'fake news' and social media[1], online misinformation has long been a global problem[2]. In fact, in many cases, the negative impact of misinformation is most starkly felt outside of North America and Western Europe. For example, in Myanmar, false information on Facebook may have facilitated genocide against the Rohingya minority group[3,4]; and in India, at least two dozen people have been killed in mob lynchings after rumours were spread on WhatsApp[5]. More generally, the worldwide nature of misinformation is perhaps most evident in the case of coronavirus disease 2019 (COVID-19). In parallel to

[1]Center for Research and Teaching in Economics (CIDE), Aguascalientes, Mexico. [2]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Centre for Decision Research and Experimental Economics, University of Nottingham, Nottingham, UK. [4]Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Google, New York, NY, USA. [6]Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [7]Rutgers University, New Brunswick, NJ, USA. [8]Department of Philosophy, Macquarie University, Sydney, New South Wales, Australia. [9]Hill/Levene Schools of Business, University of Regina, Regina, Saskatchewan, Canada. [10]Department of Psychology, University of Regina, Regina, Saskatchewan, Canada. [11]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. e-mail: grpennycook@gmail.com; drand@mit.edu

 https://doi.org/10.1038/s41562-023-01641-6

the actual pandemic, an 'infodemic' of misinformation and conspiracy theories about COVID-19 has spread around the globe[6–13], espousing false cures[14,15], questioning effective mitigation strategies (for example, regarding masks)[11,16] and promoting vaccine hesitancy[10].

Given the global reach of online misinformation, it is important to study it in a global context. There are numerous reasons to expect that the individual differences that predict susceptibility to misinformation, and the effectiveness of anti-misinformation interventions, may vary meaningfully across countries. Beyond basic issues related to generalizability from W.E.I.R.D. (Western, Educated, Industrialized, Rich and Democratic) cultures[17], the context of misinformation—and online misinformation in particular—brings unique reasons to expect variation. For example, there is a long-standing tradition of a relatively free and open press in W.E.I.R D. countries, which may lead to different attitudes, and baseline levels of credulousness, towards news than in other parts of the world[18]. W.E.I.R.D. countries also have a longer history of use of digital devices, the internet, and social media than much of the rest of the world, bringing with it a greater average level of digital literacy[19,20]. Furthermore, social media is used differently in different parts of the world. For example, while newsfeed-based platforms like Facebook are dominant in such countries, messaging platforms like WhatsApp are dominant in many other parts of the world[21]. Moreover, cultural attitudes towards accuracy, and thus the extent to which people value accuracy versus other motives when deciding what to share online, may also vary cross-culturally. Therefore, it is of great scientific importance to examine how the psychology of misinformation varies across cultures, and what patterns are consistently observed. Furthermore, from a practical perspective, social media companies—whose user bases span the globe—are understandably reluctant to implement interventions that have not been shown to have cross-cultural effectiveness.

In this Article, we shed new light on the psychology of online misinformation globally with a large-scale experiment fielded simultaneously in 16 countries across six continents (total $N = 34{,}286$; 676,605 observations). We investigate who believes and shares misinformation, and we evaluate three anti-misinformation interventions.

A major challenge for cross-cultural studies of misinformation is that each country presents a different cultural context with a unique media environment and news cycle. Thus, it is typically necessary to use different content for each country, which presents a challenge when trying to compare across countries. However, COVID-19 provided a unique opportunity in this context as it allowed us to construct a set of true and false statements that were of global relevance. In total, we selected 30 false and 15 true headlines about COVID-19 (for a full list of the headlines used in our experiment, see Supplementary Section 1). While each headline will not have the exact same level of relevance and familiarity across all countries, we aimed to create a broadly relevant headline set by compiling them from various sources including the World Health Organization's list of COVID-19 myths and fact-checking websites from several different countries. Furthermore, although (mis) information exists along a continuum of accuracy[22], for tractability we focus here on the dichotomy between clearly true and clearly false statements—while also noting that being exposed to clearly false claims increases subsequent belief just as much as exposure to more plausible false claims[23].

To evaluate who believes misinformation online and what to do about it, we specifically recruited convenience samples of social media users in each country, quota-matched to the national distribution of age and sex within each country. Although our participants were not truly representative of the populations of their respective countries (for example, our participants tended to be more educated than the general population in some countries), our samples were well calibrated to national estimates of four cultural value items from the World Values Survey in most countries (for details of the recruitment process and sample demographics, see Supplementary Section 2). Furthermore,

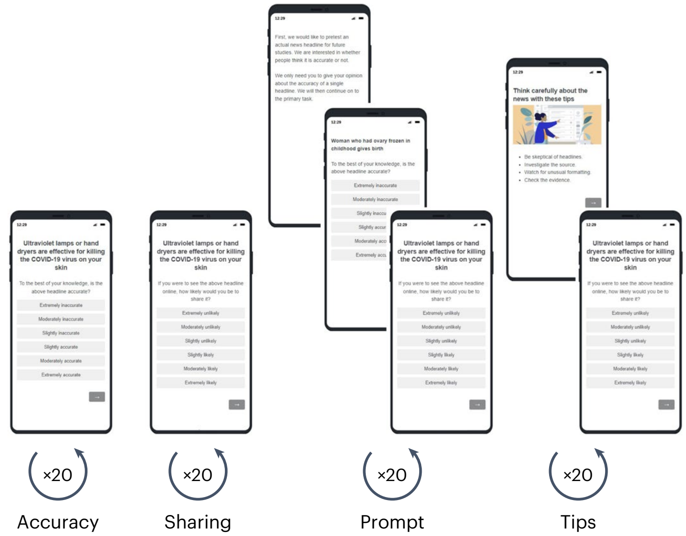

Fig. 1 | Visualization of the four experimental conditions. In the Accuracy and Sharing conditions, participants saw the same screen 20 times, with a different headline each time. In the Tips and Prompt conditions, participants first saw one or two screens implementing the treatment, and then advanced to complete the same sharing-intention screen as in the Sharing condition 20 times with different headlines each time.

data on the population of social media users (as opposed to the general population), and the relative use of different social media platforms, were not available in many countries. Thus, we do not know how closely representative our sample is of the relevant social media users in each country; although we note that the messaging application WhatsApp is included as one of the qualifying social media platforms for eligibility in our study, and WhatsApp usage is widespread in much of the world.

While misinformation comes in many forms, we follow most previous work in this area (see refs. 24,25) and focus on the belief in, and sharing of, news. We particularly focus on news headlines, rather than full articles, because on social media news is largely consumed by reading headlines without clicking through to read the full article. Specifically, we presented each participant with 10 true and 10 false news headlines about COVID-19, randomly sampled from the larger set of 45 headlines (of which 30 were false).

Each participant was also randomized into one of four experimental conditions (the Accuracy, Sharing, Prompt and Tips conditions, discussed in detail below) that varied in terms of what participants were asked about for each headline, and what (if any) interventions were applied before the headline evaluation task (Fig. 1). All analyses were pre-registered except where noted; for full survey materials, data links and our pre-registration, see Supplementary Section 1.

Who believes misinformation?

First, we test predictions generated by several theories regarding susceptibility to belief in misinformation. To do so, we examine predictors of participants' ability to identify true versus false headlines when judging their accuracy.

In the Accuracy condition, participants were asked to rate the accuracy of each headline on a scale from 1 (extremely inaccurate) to 6 (extremely accurate). Beginning with overall descriptive statistics, we find that, while participants rated true headlines as much more accurate than false headlines in every country on average, there was marked variation across countries in average truth discernment (overall accuracy of participants' judgements, computed as average ratings for true minus average ratings for false) (Fig. 2). Interestingly, this variation was largely driven by variability in the perceived accuracy of false news: On the two extremes, participants in India believed false

 https://doi.org/10.1038/s41562-023-01641-6

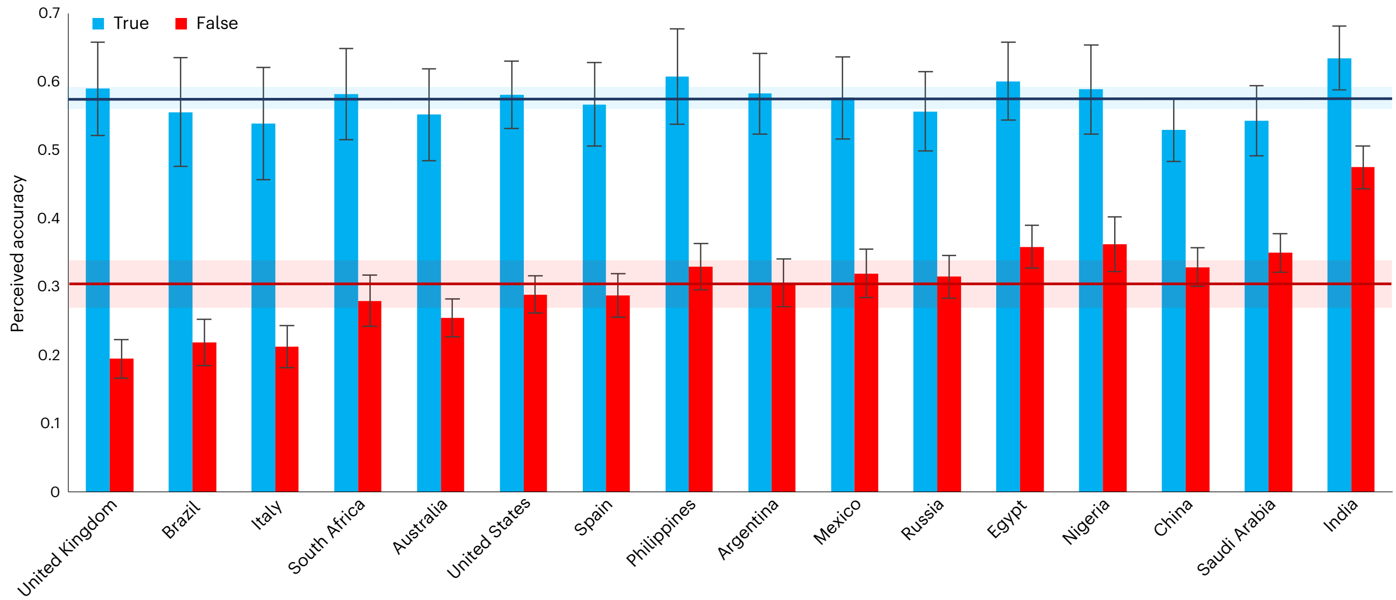

Fig. 2 | True headlines are believed more than false headlines. Bars show average accuracy ratings for true (blue) and false (red) headlines by country (error bars indicate 95% CIs), sorted by average truth discernment. Horizontal lines show meta-analytic mean estimates with 95% CIs. $n_{\text{Participants}} = 8{,}527$; $n_{\text{Ratings}} = 167{,}725$.

claims more than twice as much as participants in the United Kingdom. Conversely, there was comparatively little variability across countries in the perceived accuracy of true news. We conducted exploratory country-level analyses of the relationship between truth discernment and economic variables (inequality and gross domestic product (GDP)), cultural variables (individualism versus collectivism, and power distance), and institutional variables (corruption, freedom and human development). We find that participants from countries that are more individualistic, have more open political systems, and have lower power distance scores are significantly better at telling true headlines from false headlines (that is, have higher average truth discernment) (for details, see Supplementary Section 3.1).

What individual differences, then, predict believing misinformation? And how robust are these associations across countries? For each of 20 individual differences, we run a separate rating-level linear regression for each country, predicting perceived accuracy on the basis of the headline's objective veracity, the individual difference measure (z-scored) and their interaction, using two-way robust standard errors clustered on subject and headline. We also include demographic controls for age, sex, education and socio-economic status (and their interactions with headline veracity, as well as quadratic terms of age and socio-economic status) in these correlational analyses; we also note that excluding demographic controls from the correlational analyses has little impact on the results (Supplementary Section 3.3). We then determine the overall association, and the extent of variation across countries, using a random-effects meta-analysis. We focus on the interaction between headline veracity and each individual difference, which indicates the association between the individual difference measure and truth discernment.

Our pre-registration specified this linear regression specification and random-effects meta-analytic approach, as well as this set of demographic control variables. However, in our pre-registration, the only individual difference we indicated we would explore was performance on the Cognitive Reflection Test (CRT). The other 19 individual differences we investigate were not pre-registered, and thus those analyses should be considered post hoc.

The results are summarized in Fig. 3 (for forest plots of each individual difference, see Supplementary Fig. 5). All results shown in Fig. 3 and discussed in the associated text are robust to correcting for having conducted 20 multiple comparisons, using either the Bonferroni or Holm–Bonferroni correction methods (we did not pre-register that we would correct for multiple comparisons in these analyses, but did so as a robustness check). Post hoc analyses considering non-linear relationships do not qualitatively change the conclusions derived from the linear models reported in Fig. 3, while also revealing that more extreme responses for most Likert scale measures are associated with better truth discernment (Supplementary Section 3).

One theoretical perspective rooted in cognitive science argues that people believe misinformation when they fail to engage in analytic thinking and instead rely on their intuitions[25,26]. To test the predictions of this account, we measure self-reported preference for analytic thinking, as well as objective performance on the CRT (a set of questions with intuitively compelling but incorrect answers that is widely used to measure analytic thinking)[27]. We find a remarkably robust positive association between analytic thinking and truth discernment (Fig. 3): In every country, participants with a more analytic cognitive style were better able to discern truth from falsehood (interaction with headline veracity: self-report, meta-analytic $b = 0.037$, $z = 10.559$, $P < 0.001$, confidence interval (CI) 0.030 to 0.044; CRT, meta-analytic $b = 0.034$, $z = 9.078$, $P < 0.001$, CI 0.026 to 0.041). This result shows that robust findings from the United States context[25] generalize broadly, and emphasizes the important role of analytic thinking in truth discernment[28]. Relatedly, participants who passed more attention checks were better at telling truth from falsehood in all countries (meta-analytic $b = 0.040$, $z = 8.371$, $P < 0.001$, CI 0.030 to 0.049), as were, to a lesser extent, participants with college degrees (meta-analytic $b = 0.016$, $z = 4.353$, $P < 0.001$, CI 0.009 to 0.023; recall that all other reported associations control for demographics including education and thus emerge above and beyond this education-based association).

A more social psychological perspective emphasizes the importance of motivation in misinformation detection. While accuracy motives could drive people towards truth discernment, other motives (for example, the desire to denigrate counter-partisans[29,30], or social motivations more generally[31]) may support false beliefs. To explore the connection between accuracy motives and truth discernment, we examine two accuracy-related motivational measures. In all countries, truth discernment was higher for participants who reported placing more importance on accuracy in the context of social media sharing

 https://doi.org/10.1038/s41562-023-01641-6

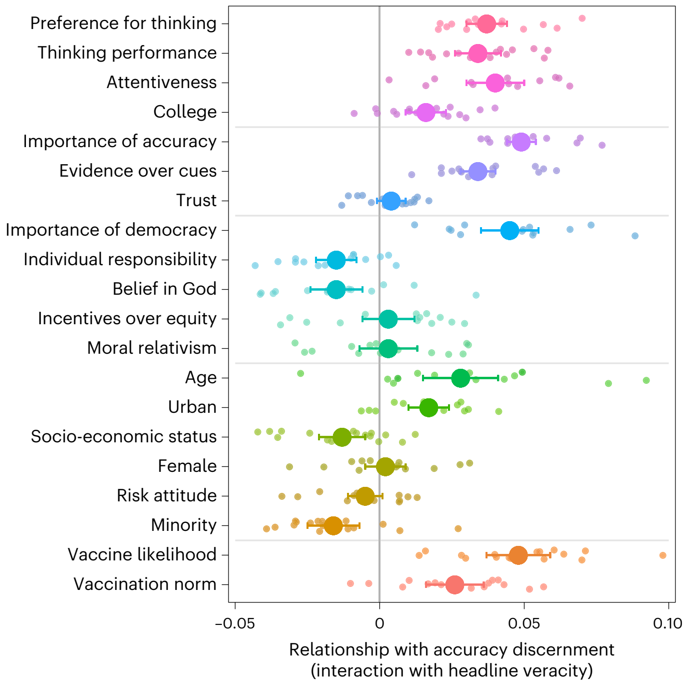

Fig. 3 | Consistent cross-cultural evidence that truth discernment is associated with analytic thinking, accuracy motivations and ideology. For each individual difference measure, the coefficient of the interaction between headline veracity and the z-scored individual difference when predicting perceived accuracy is shown. Thus, the x axis indicates the percentage point increase in truth discernment associated with a one-standard-deviation increase in the individual difference measure. The meta-analytic mean estimate and 95% CI are indicated by the large dot and error bars; the smaller dots show the mean estimate for each country. A separate model was run for each individual difference, including controls for age, sex, education and socio-economic status. For estimates labelled by country, see Supplementary Fig. 5. Estimates are based on $n = 8{,}527$ participants and 167,725 ratings.

(meta-analytic $b = 0.049$, $z = 17.313$, $P < 0.001$, CI 0.044 to 0.055) and who felt political opinions should be based on evidence and arguments more than what their party says (meta-analytic $b = 0.034$, $z = 9.837$, $P < 0.001$, CI 0.028 to 0.041). This suggests a potentially important role of motivation, in addition to the more cognitive factors discussed above, in truth discernment[32]. Another social perspective involves the role of interpersonal trust: Might susceptibility to misinformation represent a more general tendency to trust others (for example, gullibility)? To gain some insight into this possibility, we ask participants about the extent to which they trust those they interact with in daily life. We find no significant relationship between generalized trust and truth discernment (meta-analytic $b = 0.004$, $z = 1.702$, $P = 0.089$, CI −0.001 to 0.008).

A third theoretical perspective rooted in political psychology implicates ideology in susceptibility to falsehoods[33–37]. We find consistent associations with participants' responses to two items from the World Values Survey regarding government policies: Valuing democracy was associated with higher truth discernment in all countries (meta-analytic $b = 0.045$, $z = 8.800$, $P < 0.001$, CI 0.035 to 0.055), and endorsement of individual responsibility over government support was associated with worse truth discernment in most countries (meta-analytic $b = -0.015$, $z = 4.630$, $P < 0.001$, CI −0.022 to −0.009). We also find that belief in God is associated with worse truth discernment in most countries (meta-analytic $b = -0.015$, $z = 3.313$, $P = 0.001$, CI −0.024 to −0.006). The results are much more mixed for personal values that do not involve government policies, where we find that believing that incomes should be more equal, as well as moral relativism, did not show consistent associations with truth discernment. For each measure, some countries showed significant negative associations while others showed significant positive associations, and the meta-analytic results were not significant (income equality: $b = 0.003$, $z = 0.606$, $P = 0.544$, CI −0.006 to 0.012; moral relativism: $b = 0.003$, $z = 0.549$, $P = 0.583$, CI −0.007 to 0.012). These findings reveal complex and subtle relationships between ideology, culture and the ability and/or willingness to correctly tell truth from falsehood.

With respect to demographics, we find that participants who are younger, live in less urban areas, have higher subjective socio-economic status (driven particularly by the highest status participants; Supplementary Section 3.3), and identify as members of ethnic minorities in their respective countries show lower truth discernment on average, while sex and willingness to take risks are not consistently associated with truth discernment (for details, see Fig. 3).

Finally, we find a robust positive association between truth discernment and COVID-19 vaccination intentions (meta-analytic $b = 0.048$, $z = 8.957$, $P < 0.001$, CI 0.037 to 0.058); interestingly, this effect was stronger for truth discernment using vaccine-related false headlines (meta-analytic $b = 0.064$, $z = 8.847$, $P < 0.001$, CI 0.050 to 0.078) than for truth discernment using non-vaccine-related false headlines (meta-analytic $b = 0.038$, $z = 8.386$, $P < 0.001$, CI 0.029 to 0.047), although both relationships were highly significant and robust across countries (Supplementary Fig. 8). We also find a weaker—but still pronounced and fairly consistently signed—positive association between truth discernment and the extent to which participants believe that others will get vaccinated (that is, their perception of the descriptive norm regarding vaccination). These observations, although only correlational, give some reason to believe that the causal link between misinformation and vaccine hesitancy demonstrated in the United States and the United Kingdom[10] may extend more broadly.

We also conducted exploratory analyses of the extent to which variation across countries in these relationships between truth discernment and individual differences were explained by variation in country-level variables. To do so, for each combination of individual difference variable and country-level variable, we conducted a multi-level model combining data from all countries and examined the three-way interaction between headline veracity, the individual difference and the country-level variable. The results, shown in detail in Supplementary Table 6c, demonstrate the broad relevance of many of the economic, cultural and institutional factors we considered. Generally speaking, cognitive sophistication, accuracy motivations and preference for democracy were more strongly linked to truth discernment in countries that were less collectivist and corrupt, and lower on power distance. Accuracy motives and preference for democracy were also more strongly linked to truth discernment in countries that had higher GDP and human development scores and more open political systems. Support for individual responsibility over government support was more strongly negatively related to truth discernment in countries with more economic inequality. And the associations between truth discernment and the other three ideological variables were all significantly moderated in varying ways by all of the country-level variables except for economic inequality.

Accuracy judgements versus social media sharing

We now turn our attention to the sharing of misinformation on social media. Because exposure to misinformation increases belief in[38–40]—and perceived ethicality of[41]—falsehoods, understanding why people share misinformation, and how to reduce that sharing, is of great importance. As a result, there is substantial pressure on social media companies to reduce the sharing of misinformation online.

We begin by asking how tightly sharing intentions are linked to accuracy judgements. To do so, we use the Sharing condition where, instead of rating accuracy, participants indicate how likely they would be to share each headline on social media. We then compare the level

https://doi.org/10.1038/s41562-023-01641-6

of truth discernment in the Accuracy condition versus sharing discernment in the Sharing condition (where sharing discernment is defined as average sharing intentions of true headlines minus average sharing of false headlines).

In all countries, the difference between true and false headlines was greater for accuracy judgements than sharing intentions—that is, people were less discerning when deciding what to share than they were in judging accuracy (Fig. 4). Most importantly, people in the Sharing condition indicated an intention to share false headlines to a greater degree than people in the Accuracy condition believed the false headlines to be accurate (Supplementary Fig. 9). This suggests that people sometimes share false headlines that they would be able to identify as inaccurate if asked to evaluate the headline's veracity[42,43]. Individual difference predictors of sharing discernment are similar to what was observed in Fig. 3 for truth discernment (Supplementary Fig. 10); for country-level predictors of the disconnect between accuracy and sharing discernment, see Supplementary Section 3.1. This disconnect between accuracy judgements and sharing intentions is particularly notable given that, when explicitly asked at the end of the study, a large majority of participants in all countries said that accuracy was very or extremely important to them when deciding what to share online (Fig. 4 inset; for by-country breakdown, see Supplementary Fig. 2).

Do accuracy prompts increase information sharing quality?

What explains the disconnect between accuracy and sharing demonstrated in Fig. 4? Numerous factors may contribute to the sharing of false headlines one could identify as accurate, including anxiety[44], emotionality[45], distrust in science[46], the need for chaos[47], and partisanship[42]. Another possibility is that people share news they know to be false in an effort to correct or mock it; however, Twitter data suggest that this kind of behaviour is comparatively rare[48].

Here we focus on the role of inattention. Recent work has posited that mere inattention to accuracy—as opposed to confusion about veracity or purposeful sharing of falsehoods—is an important driver of the sharing of falsehoods[42,43,49,50]. If so, then simply shifting participants' attention to the concept of accuracy, without providing any additional information about the truth value of the headlines, should improve sharing discernment.

To test this prediction, we compare baseline sharing intentions in the Sharing condition with sharing intentions after receiving an accuracy prompt[51]. Specifically, participants randomly assigned to the Prompt condition began the task by being prompted to rate the accuracy of a single non-COVID-related news headline. They then completed the same sharing task as participants in the Sharing condition, but with the concept of accuracy having been brought to mind by the prompt. Thus, to the extent that inattention to accuracy is a driver of misinformation sharing, we would expect participants in the Prompt condition to be more discerning in their sharing relative to participants in the Sharing condition[43,50] (that is, the Sharing condition acts as the control condition against which the Prompt condition is compared; because participants are randomized to conditions, our analyses of these experiment effects do not include demographic controls).

As predicted, we found that the Prompt condition increased sharing discernment relative to the baseline Sharing condition (meta-analytic estimate, $b = 0.171$, $z = 4.606$, $P < 0.001$, CI 0.098 to 0.244; Fig. 5a), primarily by reducing sharing intentions for false headlines (Supplementary Fig. 11). There was significant variation across countries in the magnitude of this effect ($\chi^2 = 58.57$, $P < 0.001$; I^2 (%) = 0.744, CI 0.315 to 0.867), in a manner that is consistent with the underlying theory behind accuracy prompts[51]: If the prompt is effective because it closes the gap between accuracy judgements and sharing intentions, the intervention should be most effective for countries with the largest difference in baseline discernment for accuracy versus sharing.

Consistent with this prediction, the magnitude of the prompt effect across countries is strongly positively correlated ($r(14) = 0.762$, $P < 0.001$, CI 0.428 to 0.913; Fig. 5b) with the disconnect between truth and sharing discernment (discernment in the Accuracy condition minus discernment in the Sharing condition). Thus, the prompt is most effective for countries where people at baseline are least attentive to accuracy when deciding what to share. Relatedly, exploratory analyses indicate that the magnitude of the prompt effect was significantly larger in countries that had higher GDP, open political systems and human development scores, and were less collectivist and corrupt and lower on power distance—precisely because these countries had a larger gap between accuracy and sharing discernment; for details, see Supplementary Table 6b.

We also find an analogous relationship when examining variation in the effect of the prompt across headlines: The less accurate a headline seems (based on ratings from the Accuracy condition), the more the prompt reduces sharing of that headline relative to the baseline Sharing condition ($r(43) = 0.908$, $P < 0.001$, CI 0.838 to 0.949; Fig. 5c).

Together, these results support the hypothesized mechanism whereby the prompt improves sharing quality by shifting attention to accuracy. Together with a field experiment conducted with mostly users from the United States[43], our findings suggest that platforms could reduce the spread of certain forms of misinformation in many parts of the world by nudging users to attend to accuracy. Furthermore, we find little evidence of any individual differences that robustly moderate the treatment effect (Supplementary Fig. 12), suggesting that the intervention may be widely effective across individuals (even if there is variability across countries). These results also demonstrate the boundary conditions of the accuracy prompt approach: Shifting attention to accuracy will only reduce the sharing of misinformation in so much as users are (1) less discerning when deciding what to share than when judging accuracy (which varies across countries, see Figs. 4 and 5b) and (2) able to identify a given claim's veracity when judging accuracy (which varies across claims, see Fig. 5c, and countries, see Fig. 2).

Can minimal digital literacy tips improve sharing?

We also evaluate the effectiveness of a simple digital literacy intervention for improving sharing discernment relative to the baseline Sharing condition. Immediately before completing the sharing task, participants in the Tips condition were encouraged to think critically about the news and shown a set of four simple digital literacy tips (excerpted from an intervention developed and deployed by Facebook)[49,52]. As expected, sharing discernment was higher in the Tips condition compared with the baseline Sharing condition (meta-analytic estimate, $b = 0.076$, $z = 4.302$, $P < 0.001$, CI 0.041 to 0.110; Fig. 5d). Although this effect was smaller than the accuracy prompt effect, the magnitude of the Tips effect (in contrast to the Prompt) did not significantly vary across countries ($\chi^2 = 14.54$, $P = 0.485$; I^2 (%) = 0.000, CI 0.000 to 0.437); and accordingly, exploratory analyses found that the Tips effect was not significantly moderated by any of the country-level variables we considered (Supplementary Table 6b). Furthermore, the Tips effect was not significantly moderated by any of the individual differences we considered (Supplementary Fig. 11b).

After the sharing task in both the Prompt and Tips conditions, we explained to participants that the intervention they received at the beginning of the study was designed to help them share more accurate information. We then asked how helpful they thought the intervention was, and how positively versus negatively they felt about it. Interestingly, the tips were rated as substantially more helpful than the prompt in all countries (meta-analytic estimate: $b = 0.355$, $z = 12.811$, $P < 0.001$, CI 0.301 to 0.409; Supplementary Section 3.8)—despite the fact that the prompt was on average twice as effective as the tips in increasing sharing discernment. This highlights the limitations of simply asking people which intervention is more effective (as technology

https://doi.org/10.1038/s41562-023-01641-6

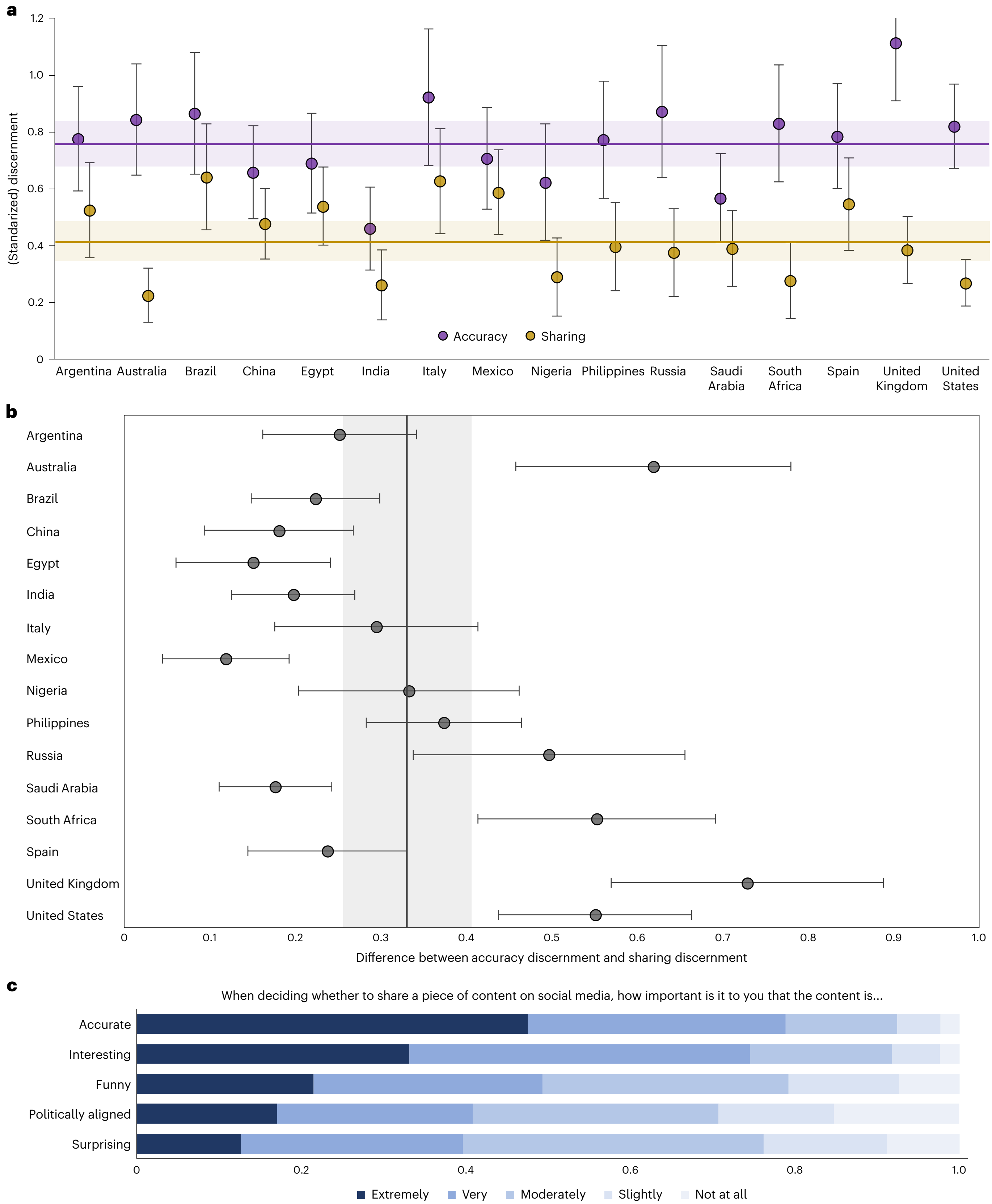

Fig. 4 | Sharing intentions are less discerning than accuracy judgements, even though people consistently rate accuracy as important when deciding what to share. a, Standardized discernment (mean value for true minus mean value for false; *z*-scored within outcome type) by country for accuracy judgements in the Accuracy condition (dots in purple) and sharing intentions in the Sharing condition (dots in gold). Error bars indicate 95% CI. Horizontal lines indicate meta-analytic mean estimates and 95% CI. $n_{\mathrm{Participants}} = 17{,}158$; $n_{\mathrm{Ratings}} = 338{,}236$. **b**, Difference between sharing and accuracy discernment by country, with meta-analytic mean difference shown with a bold vertical line and 95% CI. $n_{\mathrm{Participants}} = 17{,}158$; $n_{\mathrm{Ratings}} = 338{,}236$. **c**, Self-report importance placed on accuracy, interestingness, funniness, political alignment and surprisingness when deciding what to share online, averaged across participants. $n = 32{,}761$. For distributions by country, see Supplementary Fig. 2.

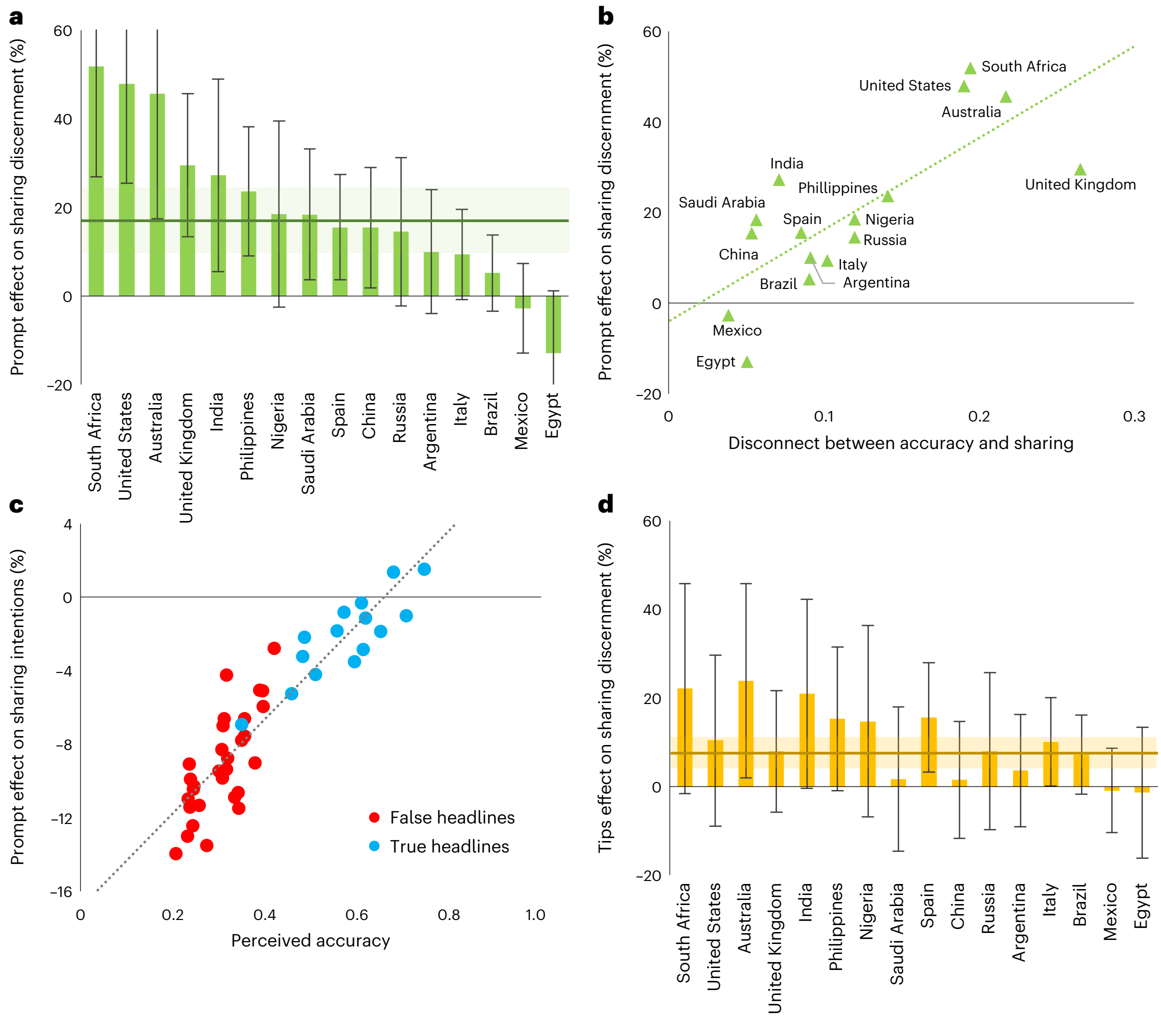

Fig. 5 | Simple interventions can improve the quality of social media sharing intentions. a, Mean per cent change in sharing discernment caused by the accuracy prompt intervention relative to the baseline (mean discernment in Prompt condition minus mean discernment in Sharing condition, all divided by mean discernment in Sharing condition). Error bars indicate 95% CIs; horizontal lines indicate meta-analytic mean estimate and 95% CI. **b**, Variation across countries in the size of the prompt effect is largely explained by variation across countries in the magnitude of the disconnect between accuracy judgements and sharing intentions. Shown on the *x* axis is discernment in the Accuracy condition minus discernment in the Sharing condition; shown on the *y* axis is discernment in the Prompt condition minus discernment in the Sharing condition. **c**, Variation across headlines in the effect of the accuracy prompt is largely explained by how accurate participants perceive the headline to be. Shown on the *x* axis is the average perceived accuracy rating from the Accuracy condition (collapsing across countries; for pre-registered by-country analysis, see Supplementary Section 3.9). Shown on the *y* axis is average sharing intention in the Prompt condition minus average sharing intention in the Sharing condition (collapsing across countries). **d**, Mean per cent change in sharing discernment caused by the digital literacy tips intervention relative to the baseline (mean discernment in Tips condition minus mean discernment in Sharing condition, all divided by mean discernment in Sharing condition). Error bars indicate 95% CIs; horizontal lines indicate meta-analytic mean estimate and 95% CI. For **a**–**d**, $n_{\text{Participants}} = 34{,}286$; $n_{\text{Ratings}} = 676{,}605$.

companies often do), and emphasizes the importance of directly assessing the effectiveness of interventions[53]. From a practical perspective, it is also important that, for both interventions, the large majority of participants in all countries were either neutral or positive (84.2% neutral or positive ratings for prompt, 97.0% for tips; Supplementary Section 3.8). Thus, it seems likely that there would be little public resistance to either intervention should they be adopted by social media platforms or policy makers.

Can layperson accuracy ratings help identify misinformation?

Finally, we turn from the judgements of individuals to the judgements of groups. The sheer volume of content posted online every day poses a major challenge for efforts to combat misinformation. Professional fact-checking is a time-consuming process and requires specialized training. As a result, the fraction of content that can be checked by professionals is minuscule. This is particularly true in countries that do not have a robust press and tradition of professional fact-checking. Thus, although professional fact-checks are extremely useful when they are available, employing them at scale is challenging.

Here we ask whether layperson accuracy judgements can be leveraged to help identify misinformation at scale[22,54,55]. From a practical perspective, the answer involves not just whether the ratings of the crowd are well calibrated to ground truth, but also whether a high level of agreement with ground truth can be reached with a relatively small crowd. Thus, we ask how effectively average ratings of participants from each country can be used to identify true versus false COVID-19 statements as a function of the number of participant ratings per headline (that is, the size of the 'crowd'). See Supplementary Section 1.4 for details of the sampling procedure used to determine the area under the

https://doi.org/10.1038/s41562-023-01641-6

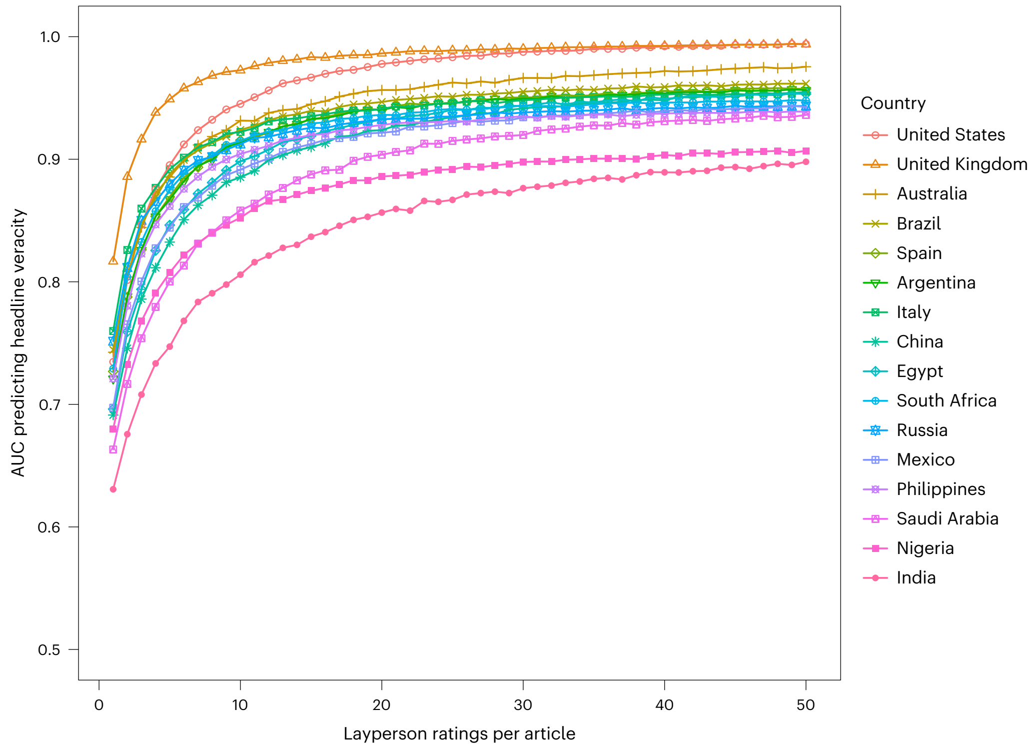

Fig. 6 | Ratings from even small groups of laypeople can reliably distinguish true from false headlines. AUC when predicting headline veracity using the average rating of a crowd of *k* layperson respondents, for each country. AUC can be interpreted as the probability that, when randomly selecting one true headline and one false headline, the true headline will have a higher accuracy rating than the false headline. For details, see Supplementary Section 1.4.

curve (AUC) for different crowd sizes, which was not pre-registered but is identical to the procedure used in previous work[22,54,55].

We find that, in almost all countries, as few as 15 ratings per headline are enough to differentiate true from false headlines over 90% of the time; and in all countries, 10 ratings per headline enabled differentiation over 80% of the time (Fig. 6). This demonstrates that the potential for crowdsourcing to help identify misinformation is not restricted to the United States[22,54,55] (just as it is not restricted to highly educated subjects; Supplementary Section 3.10), despite there being some variability across countries.

Conclusion

Misinformation is a global problem that requires evidence-based solutions that are not idiosyncratic to particular cultural contexts. In the large cross-cultural experiment reported here, we find some reason for optimism about such efforts: Across 16 countries on all 6 inhabited continents, we find striking regularities in both the underlying psychology of misinformation and the effectiveness of interventions to combat it.

Although average levels of belief in falsehoods did vary substantially across countries, we found consistent evidence that analytic thinking, accuracy motivations and support for democracy were associated with a greater ability to discern truth from falsehood, as well as fairly consistent evidence that endorsement of individual responsibility over government support and belief in God were associated with worse truth discernment. These regularities emphasize the joint importance of cognitive and social factors, and suggest that a common psychology may underlie susceptibility to COVID-19 misinformation across cultural contexts. They also help identify individuals who are most at risk of falling prey to misinformation, and thus would benefit most from anti-misinformation interventions.

Our results also highlight the challenges that misinformation poses for social media platforms in particular. In all countries, we found (at least some) evidence that people share news they would be able to identify as false if asked. An important implication of this disconnect between accuracy and sharing is that education campaigns and media literacy training aimed at improving the ability to identify falsehoods—although certainly positive—are unlikely to be sufficient on their own to stop the spread of misinformation. It is also critical to address the features of social media and society that may distract or disinhibit people from prioritizing truth.

Our observation that the effectiveness of anti-misinformation interventions developed in the United States generalizes broadly across countries is particularly encouraging. Our results suggest that shifting users' attention to the concept of accuracy may be effective at reducing the sharing of misinformation, particularly in countries where there is a substantial disconnect between accuracy judgements and sharing intentions. On the other hand, accuracy prompts are unlikely to be helpful in countries where this disconnect is small (either because accuracy discernment is low or sharing discernment is already comparatively high), or for inaccurate claims that are widely believed.

Our results also suggest that digital literacy tips may be widely helpful. Our study may in fact underestimate the effect of literacy tips, as the tips we provided were quite minimal and some were not useful in the restricted context of our survey experiment (for example, instructions to pay attention to the source, as source information was not provided). Future work should investigate the efficacy of more detailed literacy interventions in richer settings.

The ability to identify false claims using the aggregated accuracy ratings of small groups of laypeople suggests that the wisdom of crowds may be a potent tool for helping to extend the reach of fact-checking (for example, for informing warning labels or ranking algorithm demotion). Of course, an important challenge for the crowdsourcing approach is the possibility of misuse. For example, bad actors can execute coordinated attacks where they inappropriately flag accurate content that they disagree with. Approaches for helping to prevent misuse include systems where raters are given randomly selected pieces of content to rate (rather than being able to choose which pieces of content to evaluate), or where raters have to earn (and maintain) a reputation for high-quality ratings in order for their ratings to be counted. For further discussion of crowdsourced content evaluation, see ref. 56.

Our general observation of cross-cultural intervention effectiveness resonates with recent findings from smaller-scope cross-cultural projects that found, for example, that digital literacy tips improved accuracy discernment in the United States and India[52] and that fact-checks reduced belief in false claims in Argentina, Nigeria, South Africa and the United Kingdom[57]. In addition to highlighting specific interventions that appear promising, these results more broadly suggest that interventions designed and tested using W.E.I.R.D. populations, so long as they are rooted in basic psychological mechanisms, may be able to transcend cultural differences and help combat misinformation (as well as be seen positively by users) around the globe. When considering interventions to combat misinformation, it is also important to bear in mind that no single solution is a panacea that will solve the problem on its own. Instead, making progress against misinformation requires expanding the toolkit of successful approaches and applying them in combination[58].

A limitation of our study, of course, is that we used convenience samples in every country. Although they were quota-matched on the basis of age and sex, these samples were not fully representative of the general populations of their respective countries, or of social media users in their respective countries (given a lack of good data on the population of users of each different social media platform in many countries, it is extremely hard to assess how representative our sample was of the national distribution of relevant social media users). Of particular potential concern, the education levels sampled were substantially higher than the national average in some countries. Encouragingly, however, we did not find that education (or any of the other individual differences we measured) moderated the effects of the interventions we evaluated. Thus, there is some reason to believe that the results we observe will generalize to more representative samples. Furthermore, despite the non-representativeness, the results presented here at the very least demonstrate that patterns observed in previous work are not unique to the United States and Western Europe. Nonetheless, it is important for future research to investigate the issues we explore in this paper using other, more representative, samples (for example, non-internet panels)[52], and panels that attempt to mitigate issues around what type of respondent opts into completing surveys.

Our stimulus set presents another important set of limitations. To generate a headline set that was as globally salient as possible, we used global resources to source the headlines (for example, the World Health Organization) and avoided headlines that were specific to any of the countries included in the study. Nonetheless, the level of exposure to, and thus familiarity with, any given headline in our study undoubtedly varied substantially across countries. It is possible that this variation in familiarity may have influenced our results, and future studies should investigate this issue by seeking to balance familiarity levels across countries. For example, instead of using the same set of headlines in all countries, future work could select the headlines with the highest level of social media engagement in each country. We draw some reassurance, however, from the observation that countries where familiarity with our headlines was probably highest (for example, the United States and the United Kingdom) showed some of the highest levels of truth discernment, despite familiarity being consistently linked to lower levels of discernment in prior work[38,59]. Additionally, re-analysing the data from the Prompt treatments in Studies 3–5 of Pennycook et al.[43] found that, across the 68 headlines used in those studies, the size of the prompt effect was strongly predicted by the headline's perceived accuracy, $b = 0.822$, $P < 0.001$, CI 0.601 to 1.044 (as we find in Fig. 5c), and not by the headline's level of familiarity, $b = 0.030$, $P = 0.789$, CI −0.192 to 0.251. This suggests that variation across countries in familiarity with a given headline may not alter the effect of the Prompt treatment.

In addition to these issues related to familiarity, another limitation of our stimulus set is that we presented only news headlines. Future work should investigate how our findings generalize to settings where full articles are available (for example, by clicking on headlines in a newsfeed), and to misinformation that comes in other forms (for example, messages, videos, memes or posts from other users that do not contain news links). Furthermore, we examined only relatively clear-cut cases of true versus false statements. Future work should investigate a broader range of misinformation, including claims that are misleading rather than outright false, as well as examples of propaganda[60] or, more generally, rumours[61]. In a similar vein, we focused on misinformation about COVID-19 and examined a specific set of 45 headlines. It is important for future work to assess how our findings generalize to other sets of COVID-19 headlines, and to misinformation topics beyond COVID-19.

Another limitation is that our measures of sharing were hypothetical. However, prior work has suggested that self-report sharing intentions show similar association patterns to actual sharing[62], and the Prompt intervention tested here has been shown to affect actual sharing in a Twitter field experiment[43]. Furthermore, the pattern of results we observe does not seem to suggest social desirability/demand effects. Such concerns would lead people to exaggerate their level of sharing discernment (for example, by under-reporting sharing intentions for false news)—yet a key finding in our data is that sharing discernment is surprisingly low in the Sharing condition. Furthermore, one would expect that the Tips condition, which explicitly instructs participants to be more discerning, would lead to more demand effects than the fairly subtle Prompt condition—yet the Prompt condition had a substantially larger effect on sharing discernment than the Tips condition. Thus, although cross-cultural social media field experiments examining these interventions on-platform are a critical direction for future work, there is good reason to expect our sharing intentions findings to extend to actual sharing.

In sum, the results reported here help move us closer to addressing misinformation on a global scale. The broadly cross-culturally consistent patterns we observe suggest that countries around the world face similar psychological factors underlying the misinformation challenge—and can be equipped with similar solutions to meet this challenge.

Methods

We showed participants 20 COVID-19-related headlines, half of them true and half of them false. Depending on the condition assigned, they were asked to rate either the level of accuracy or their likelihood of sharing such content on social media. The study was conducted in 16 countries and 9 languages (in parentheses): Argentina (Spanish), Australia (English), Brazil (Portuguese), China (Mandarin), Egypt (Arabic), India (Hindi or English), Italy (Italian), Mexico (Spanish), Nigeria (English), the Philippines (Tagalog or English), Russia (Russian), Saudi Arabia (Arabic), Spain (Spanish), United Kingdom (English), United States (English) and South Africa (English).

Participants

We pre-registered a target sample of 2,000 participants per country, recruited through Lucid Marketplace using country-specific representative quotas on age and sex. We aimed for the same sample size in each country to provide a consistent level of statistical power, rather than aiming to reflect differences across countries in population size. We also specified that participants would not be allowed to complete the study if (1) they failed either of two trivial attention checks at the study outset or (2) they reported not having any social media accounts, declared also at the study outset. In total, 54,757 participants began the survey and 20,216 reported not having any social media accounts or

https://doi.org/10.1038/s41562-023-01641-6

failed the initial attention checks and were not allowed to continue. In addition, 255 did not provide any ratings, thus leaving 34,286 respondents with at least one rating (676,605 observations in total), and 33,480 with a complete set of 20 ratings. No country had fewer than 1,928 complete responses and it took the median participant 15:42 min to complete the entire study. Mean age of the participants was 38.7 years old, and 45% were female (for details, see Supplementary Table 2).

It took 63 days to complete data collection (from 22 February to 25 April 2021). However, 68% of the sample was gathered within the first week, and 95% within the first 24 days (Supplementary Fig. 1). The remaining 5% of the observations collected since then corresponded to age and sex quotas that were particularly hard to reach in a few countries. Indeed, the resulting age and sex distribution by country closely mirrored that of their respective populations. If anything, older subgroups were under-represented in some countries, but this could in fact be closer to the representative sample one would expect if social media users were the target population (for details, see Supplementary Fig. 3).

Materials

We asked participants to complete a 15 min survey programmed in Qualtrics. This software and the rules set by the supplier prevented people from participating more than once. The base questionnaire had 71 questions, but in some countries a few questions deemed as non-essential for this project were dropped to keep the survey within the expected time for completion (which varied across countries). The questionnaire and list of headlines, as shown in the United States, can be found in Supplementary Section 1. For other countries, we recruited translators from the website Upwork and asked them to translate all materials into their local language; for English-speaking countries, we asked them to localize terms to sound more natural. Once we had the translated documents, we recruited another translator from the same website and asked them to back-translate the materials (they were not aware that the original language of the documents was English). Back-translated documents were then reviewed by a native English-speaking author of this manuscript (D.G.R.), and in case of discrepancies another author (A.A.A.) coordinated further rounds of review with translators, or back-translators, until a satisfactory outcome was reached. Translators also tested the final version of the programmed survey before deployment. Materials are available at https://osf.io/g65qu/. We used R 3.6.1, RStudio 2022.07.0+548 and Stata 15 for data analysis.

Procedure

Participants were randomly allocated to one of four conditions: Accuracy, Sharing, Prompt or Tips. Eligible participants were then shown a set of 10 false and 10 true COVID-related headlines (one at a time and randomly sampled from a list of 30 false and 15 true headlines). We asked them to assess either the accuracy of a headline in the Accuracy condition (‘To the best of your knowledge, is the above headline accurate?’; 6-point Likert scale) or, for the other three conditions, the likelihood of sharing a given headline (‘If you were to see the above headline online, how likely would you be to share it?’; 6-point Likert scale). In the Sharing condition, participants were simply asked about their sharing intentions for each item. In the Prompt condition, participants were first asked to evaluate the accuracy of an unrelated headline (randomly selected from a list of four), and for the Tips condition, they were first shown four digital literacy tips, originally developed by Facebook and implemented in the United States and India[52]. Once the task was completed, participants completed a three-item CRT[27], several questions aiming to explore individual difference moderators and, for a subset of the countries, questions that will be used as part of separate projects. Finally, participants were debriefed. Specifically, we re-presented true headlines they had been shown and informed participants that these headlines were all true, and any headlines not shown were false (we did not re-present the false headlines to avoid the risk of exposure effects[38]).

Ethics

This research was deemed exempt by the MIT Committee on the Use of Humans as Experimental Subjects, #E-2982. Informed consent was obtained from all participants (see page 4 of Supplementary Materials).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are accessible through this link: https://osf.io/g65qu/.

Code availability

Code and materials are accessible through this link: https://osf.io/g65qu/.

References

1. Lazer, D. et al. The science of fake news. *Science* **9**, 1094–1096 (2018).
2. Bradshaw, S. & Howard, P. N. *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation* (Univ. Oxford, 2019).
3. Whitten-Woodring, J., Kleinberg, M. S., Thawnghmung, A. & Thitsar, M. T. Poison if you don’t know how to use it: Facebook, Democracy, and Human Rights in Myanmar. *Int. J. Press.* **25**, 407–425 (2020).
4. Mozur, P. A genocide incited on Facebook, with posts from Myanmar’s military. *The New York Times* (15 October 2018).
5. Arun, C. On WhatsApp, rumours, and lynchings. *Econ. Polit. Wkly* **54**, 7–8 (2019).
6. Khandelwal, D., Gildejeva, K. & Miller, E. Covid lies are tearing through India’s family WhatsApp groups. *Wired* (14 April 2021).
7. Lederer, E. UN chief says misinformation about COVID-19 is new enemy. *AP News* (28 March 2020).
8. Roozenbeek, J. et al. Susceptibility to misinformation about COVID-19 around the world: susceptibility to COVID misinformation. *R. Soc. Open Sci.* **7**, 201199 (2020).
9. Basch, C. H., Meleo-Erwin, Z., Fera, J., Jaime, C. & Basch, C. E. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Hum. Vaccin Immunother.* **17**, 2373–2377 (2021).
10. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K. & Larson, H. J. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav*. **5**, 337–348 (2021).
11. Pennycook, G., McPhetres, J., Bago, B. & Rand, D. G. Beliefs about COVID-19 in Canada, the United Kingdom, and the United States: a novel test of political polarization and motivated reasoning. *Personal. Soc. Psychol. Bull*. **48**, 750–765 (2022).
12. Bursztyn, L., Rao, A., Roth, C. & Yanagizawa-Drott, D. Misinformation during a pandemic. *Becker Friedman Inst. Work. Pap*. https://bfi.uchicago.edu/working-paper/2020-44/ (2020).
13. Brennen, J. S., Simon, F., Howard, P. N. & Kleis Nielsen, R. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute* https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation (2020).
14. Fleming, N. Coronavirus misinformation, and how scientists can help to fight it. *Nature* **583**, 155–156 (2020).
15. Jain, S. India’s healthcare workers are busting misinformation on WhatsApp. *The Verge* https://www.theverge.com/22535642/covid-misinformation-india-asha-whatsapp (2021).

16. Simonov, A., Sacher, S., Dube, J.-P. & Biswas, S. The persuasive effect of Fox News: non-compliance with social distancing during the COVID-19 pandemic. *Natl. Bur. Econ. Res.* https://www.nber.org/system/files/working_papers/w27237/w27237.pdf (2020).
17. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
18. Fawzi, N. et al. Concepts, causes and consequences of trust in news media—a literature review and framework. *Ann. Int. Commun. Assoc.* **45**, 154–174 (2021).
19. Mardikyan, S., Yıldız, E., Ordu, M. & Şimşek, B. Examining the global digital divide: a cross-country analysis. *Commun. IBIMA* https://doi.org/10.5171/2015.592253 (2015).
20. Sanou, B. ICT facts and figures 2016. *International Telecommunication Union* https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf (2016).
21. Kemp, S. Digital 2021: Global Overview Report. *DataReportal* https://datareportal.com/reports/digital-2021-global-overview-report (2021).
22. Allen, J., Arechar, A. A., Pennycook, G. & Rand, D. G. Scaling up fact-checking using the wisdom of crowds. *Sci. Adv.* **7**, 36 (2021).
23. Fazio, L. K., Rand, D. G. & Pennycook, G. Repetition increases perceived truth equally for plausible and implausible statements. *Psychon. Bull. Rev.* **26**, 1705–1710 (2019).
24. Pennycook, G., Binnendyk, J., Newton, C. & Rand, D. G. A practical guide to doing behavioural research on fake news and misinformation. *Collabra Psychol.* **7**, 25293 (2021).
25. Pennycook, G. & Rand, D. G. The psychology of fake news. *Trends Cogn. Sci.* **25**, 388–402 (2021).
26. Pennycook, G. & Rand, D. G. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
27. Frederick, S. Cognitive reflection and decision making. *J. Econ. Perspect.* **19**, 25–42 (2005).
28. Bago, B., Rand, D. & Pennycook, G. Fake news, fast and slow: deliberation reduces belief in false (but not true) news headlines. *J. Exp. Psychol. Gen.* **149**, 1608–1613 (2020).
29. Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A. & Petersen, M. B. Partisan polarization is the primary psychological motivation behind 'fake news' sharing on Twitter. *Am. Polit. Sci. Rev.* **115**, 999–1015 (2021).
30. Arceneaux, K. et al. Some people just want to watch the world burn: the prevalence, psychology and politics of the 'Need for Chaos'. *Phil. Trans. R. Soc. B* **376**, 20200147 (2021).
31. Grant, A. M. & Shandell, M. S. Social motivation at work: the organizational psychology of effort for, against, and with others. *Annu. Rev. Psychol.* **73**, 301–326 (2022).
32. Rathje, S., Roozenbeek, J., Van Bavel, J. J. & van der Linden, S. Accuracy and social motivations shape judgements of (mis) information. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-023-01540-w (2023).
33. Pretus, C. et al. The role of political devotion in sharing partisan misinformation. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/7k9gx (2021).
34. Van Bavel, J. J. & Pereira, A. The partisan brain: an identity-based model of political belief. *Trends Cogn. Sci.* **22**, 213–224 (2018).
35. Jost, J. T. & Krochik, M. Ideological differences in epistemic motivation: implications for attitude structure, depth of information processing, susceptibility to persuasion, and stereotyping. *Adv. Motiv. Sci.* **1**, 181–231 (2014).
36. Rutjens, B. T., Sutton, R. M. & van der Lee, R. Not all skepticism is equal: exploring the ideological antecedents of science acceptance and rejection. *Personal. Soc. Psychol. Bull.* **44**, 384–405 (2018).
37. Campbell, T. H. & Kay, A. C. Solution aversion: on the relation between ideology and motivated disbelief. *J. Pers. Soc. Psychol.* **107**, 809–824 (2014).
38. Pennycook, G., Cannon, T. D. & Rand, D. G. Prior exposure increases perceived accuracy of fake news. *J. Exp. Psychol. Gen.* **147**, 1865–1880 (2018).
39. Smelter, T. J. & Calvillo, D. P. Pictures and repeated exposure increase perceived accuracy of news headlines. *Appl. Cogn. Psychol.* **34**, 1061–1071 (2020).
40. Calvillo, D. P. & Smelter, T. J. An initial accuracy focus reduces the effect of prior exposure on perceived accuracy of news headlines. *Cogn. Res. Princ. Implic.* **5**, 55 (2020).
41. Effron, D. A. & Raj, M. Misinformation and morality: encountering fake-news headlines makes them seem less unethical to publish and share. *Psychol. Sci.* **31**, 75–87 (2020).
42. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. & Rand, D. G. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy nudge intervention. *Psychol. Sci.* **31**, 770–780 (2020).
43. Pennycook, G. et al. Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).
44. Freiling, I., Krause, N. M., Scheufele, D. A. & Brossard, D. Believing and sharing misinformation, fact-checks, and accurate information on social media: the role of anxiety during COVID-19. *N. Media Soc.* **25**, 141–162 (2023).
45. Li, M., Chen, Z. & Rao, L.-L. Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Comput. Hum. Behav.* **133**, 107295 (2022).
46. Saling, L. L., Mallal, D., Scholer, F., Skelton, R. & Spina, D. No one is immune to misinformation: an investigation of misinformation sharing by subscribers to a fact-checking newsletter. *PLoS ONE* **16**, e0255702 (2021).
47. Petersen, M., Osmundsen, M. & Arceneaux, K. The "need for chaos" and motivations to share hostile political rumors. *Am. Polit. Sci. Rev.* https://doi.org/10.1017/S0003055422001447 (2023).
48. Mosleh, M., Martel, C., Eckles, D. & Rand, D. G. Perverse downstream consequences of debunking: being corrected by another user for posting false political news increases subsequent sharing of low qality, partisan, and toxic content in a twiter field experiment. In Proc. 2021 *CHI Conference on Human Factors in Computing Systems* (eds Kitamura, Y. et al.) 1–13 (Association for Computing Machinery, 2021).
49. Epstein, Z. et al. Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online. *Harv. Kennedy Sch. Misinformation Rev.* https://doi.org/10.37016/mr-2020-71 (2021).
50. Pennycook, G. & Rand, D. G. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nat. Commun.* **13**, 2333 (2022).
51. Pennycook, G. & Rand, D. G. Nudging social media sharing towards accuracy. *Ann. Am. Acad. Pol. Soc. Sci.* **700**, 152–164 (2022).
52. Guess, A. M. et al. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proc. Natl Acad. Sci. USA* **117**, 15536–15545(2020).
53. Pennycook, G. & Rand, D. G. The right way to fight fake news. *The New York Times* (24 March 2020).
54. Godel, W. et al. Moderating with the mob: evaluating the efficacy of real-time crowdsourced fact-checking. *J. Online Trust Saf.* https://doi.org/10.54501/jots.v1i1.15 (2021).
55. Resnick, P., Alfayez, A., Im, J. & Gilbert, E. Searching for or reviewing evidence improves crowdworkers' misinformation judgments and reduces partisan bias. *Collective Inteligence* https://doi.org/10.1177/26339137231173407 (2023).
56. Martel, C., Allen, J. N. L., Pennycook, G. & Rand, D. Crowds can effectively identify misinformation at scale. *Perspect. Psychol. Sci.* (in the press).

57. Porter, E. & Wood, T. J. The global effectiveness of fact-checking: evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proc. Natl Acad. Sci. USA* **118**, e2104235118 (2021).
58. Bak-Coleman, J. et al. Combining interventions to reduce the spread of viral misinformation. *Nat. Hum. Behav* **6**, 1372–1380 (2022).
59. Pennycook, G. & Rand, D. G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Pers.* **88**, 185–200 (2020).
60. Faris, R. M. et al. Partisanship, propaganda, and disinformation: online media and the 2016 U.S. Presidential Election. *Berkman Klein Cent. Internet Soc. Res. Pap*. https://ssrn.com/abstract=3019414 (2017).
61. Berinsky, A. Rumors and health care reform: experiments in political misinformation. *Br. J. Polit. Sci*. **47**, 241–262 (2017).
62. Mosleh, M., Pennycook, G. & Rand, D. G. Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLoS ONE* **15**, e0228882 (2020).

Acknowledgements

The authors gratefully acknowledge funding from: The MIT Sloan Latin America Office; The Sloan Foundation; The National Science Foundation (2047152); The Ethics and Governance of Artificial Intelligence Initiative of the Miami Foundation; The William and Flora Hewlett Foundation; The Reset Initiative of Luminate (part of the Omidyar Network); The John Templeton Foundation; The TDF Foundation; The Canadian Institutes of Health Research; The Social Sciences and Humanities Research Council of Canada; The Australian Research Council (DP180102384); Google. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.A.A., A.J.B., G.P. and D.G.R. conceived the research; A.A.A., J.A., A.J.B., R.C., Z.E., K.G., A.G., J.G.L., R.M.R., M.N.S., Y.Z., G.P. and D.G.R. designed the study; A.A.A. conducted the study; A.A.A., J.A. and D.G.R. analysed the data; A.A.A., G.P. and D.G.R. wrote the paper with input from J.A., A.J.B., R.C., Z.E., K.G., A.G., J.G.L., R.M.R., M.N.S. and Y.Z.

Competing interests

A.J.B., G.P. and D.G.R. received research support through gifts from Google. G.P. and D.G.R. received research support through gifts from Facebook. R.C. and A.G. were Faculty Research Fellows at Google for several months in 2022. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01641-6.

Correspondence and requests for materials should be addressed to Gordon Pennycook or David G. Rand.

Peer review information *Nature Human Behaviour* thanks Nathan Walter, Michael Hameleers and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): David G. Rand

Last updated by author(s): Feb 14, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used Qualtrics survey software (2021) to collect the data.
Data analysis	We used R 3.6.1, RStudio 2022.07.0+548, and Stata 15 for data analysis. Code available at: https://osf.io/g65qu/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability: Data are accessible through this link: https://osf.io/g65qu/.
Code availability: Code and materials are accessible through this link: https://osf.io/g65qu/.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Our findings do not apply to only one sex or gender. Participants were recruited to quota-match the national distribution of sex within each country surveyed, and sex was self-reported. We obtained consent to share anonymized data and we include disaggregated sex data in our source dataset: 24,422 participants were female (out of 54,757). Some of our analyses included a control for sex (among others).
Population characteristics	The study was conducted in 16 countries and nine languages (in parentheses): Argentina (Spanish), Australia (English), Brazil (Portuguese), China (Mandarin), Egypt (Arabic), India (Hindi or English), Italy (Italian), Mexico (Spanish), Nigeria (English), the Philippines (Tagalog or English), Russia (Russian), Saudi Arabia (Arabic), Spain (Spanish), United Kingdom (English), United States (English), and South Africa (English). Mean age of the participants was 38.7 years old, and 45% were female.
Recruitment	We pre-registered a target sample of 2,000 participants per country, recruited through Lucid Marketplace using country-specific representative quotas on age and sex. We also specified that participants would not be allowed to complete the study if (i) they failed either of two trivial attention checks at the study outset or (ii) they reported not having any social media accounts, declared also at the study outset. 54,757 participants began the survey, 20,216 reported not having any social media accounts or failed the initial attention checks and were not allowed to continue, 1,061 failed to complete the study, and 33,480 completed the study. No country had less than 1,928 complete responses and it took the median participant 15:42 minutes to complete the entire study.
Ethics oversight	This research was deemed exempt by the MIT Committee on the Use of Humans as Experimental Subjects, #E-2982.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Study design is outlined in the Methods section of the main paper and in its Supplementary Materials. This study presents quantitative experimental analyses (we showed participants 20 COVID-19-related headlines, half of them true and half of them false. Depending on the condition assigned, they were asked to rate either the level of accuracy or their likelihood of sharing such content on social media) with data obtained from 33,480 self-report questionnaires.
Research sample	Participants were recruited online using the survey provider Lucid, with quota-matched samples based on each country's national distributions of age and sex. The study was conducted in 16 countries and nine languages (N=34,286; 676,605 observations; languages available in parentheses): Argentina (Spanish), Australia (English), Brazil (Portuguese), China (Mandarin), Egypt (Arabic), India (Hindi or English), Italy (Italian), Mexico (Spanish), Nigeria (English), the Philippines (Tagalog or English), Russia (Russian), Saudi Arabia (Arabic), Spain (Spanish), United Kingdom (English), United States (English), and South Africa (English). No country had fewer than 1,928 complete responses. It took the median participant 15:42 minutes to complete the entire study. Mean age of the participants was 38.7 years old, and 45% were female.
Sampling strategy	We based our sample sizes on comparable previous work we had conducted in the United States, and determined a pre-registered target sample of 2,000 responses per country. Our sampling procedure relied on online convenience samples provided by Lucid that were quota-matched to each country's national distributions of age and sex.
Data collection	We collected data through Qualtrics (2021; a popular online survey software). Researchers were blind to the experimental conditions participants were assigned to but were not blind to the study hypotheses during data collection. Data were collected online on participants' computers or phones. Researchers were not with participants during data collection.
Timing	From February 22 to April 25, 2021.
Data exclusions	We pre-registered a target sample of 2,000 participants per country, recruited through Lucid Marketplace using country-specific representative quotas on age and sex. We also specified that participants would not be allowed to complete the study if (i) they failed either of two trivial attention checks at the study outset or (ii) they reported not having any social media accounts, declared also at the study outset. 54,757 participants began the survey, 20,216 reported not having any social media accounts or failed the initial attention checks and were not allowed to continue, 1,061 failed to complete the study, and 33,480 completed the study.

Non-participation	A total of 1,061 participants dropped out of the study partway through but no reason was given for doing so.
Randomization	Participants were automatically randomized into conditions at the study outset.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging