This manuscript has been published in Nature Human Behaviour

https://www.nature.com/articles/s41562-023-01641-6

Understanding and Combating Misinformation Across 16 Countries on Six Continents

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*]

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

Abstract

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 COVID-19 misinformation, and interventions to combat the spread of COVID-19 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 was broadly effective at improving the veracity of news that people were willing to share, as were 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 rumors were spread on WhatsApp[5]. More generally, the worldwide nature of misinformation is perhaps most evident in the case of COVID-19. In parallel to 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 (e.g., 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 those cultures, 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.

Here, 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 (see SI Section 1 for a full list of the headlines used in our experiment). 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 believe 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 (e.g. 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 (see SI Section 2 for details of the recruitment process and sample demographics). Furthermore, consistent data on the population of social media users (as opposed to the general population), and the relative use of different social media platforms, was 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 a larger set of 45 headlines (of which 30 were false).

Each participant was also randomized into 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 prior to the headline evaluation task; see Figure 1. All analyses were pre-registered except where noted; for full survey materials, data links, and our pre-registration, see SI Section 1.

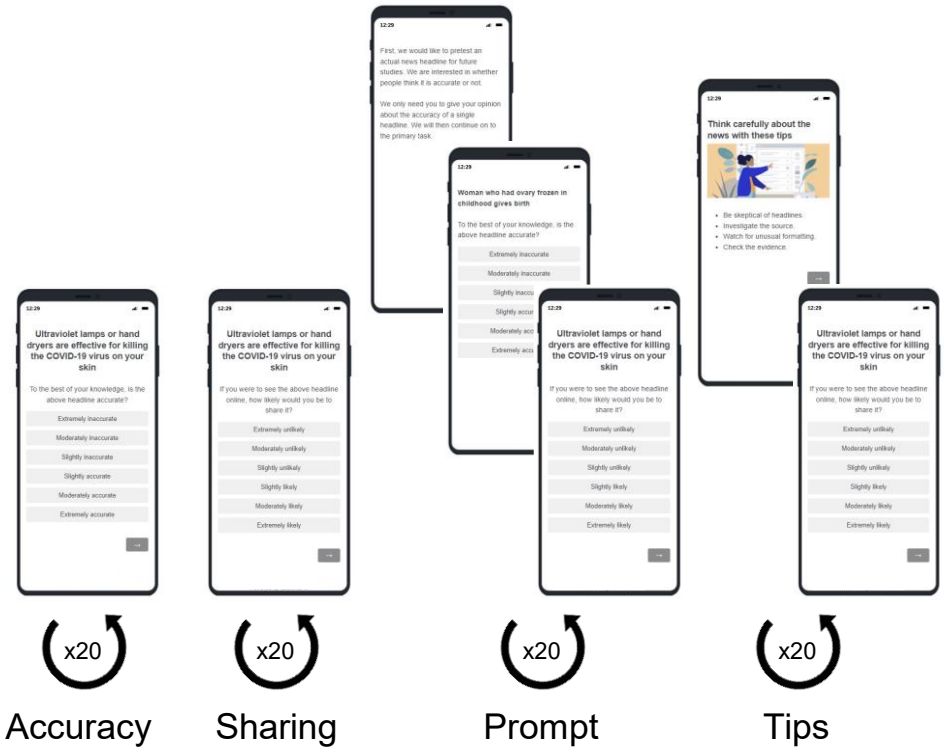


Fig. 1. Visualization of the four experimental conditions. In the Accuracy and Sharing conditions, participants saw the same screen 20 times, with a different headline each time. In the Tips and Prompt conditions, participants first saw one or two screens implementing the treatment, and then advanced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

Who believes misinformation?

First, we test predictions generated by several theories regarding susceptibility to 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' judgments, computed as average ratings for true minus average ratings for false); see Figure 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 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, GDP), cultural variables (individualism versus collectivism, power distance), and institutional variables (corruption, freedom, 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 (i.e. have higher average accuracy discernment); see SI Section 3.1 for details.

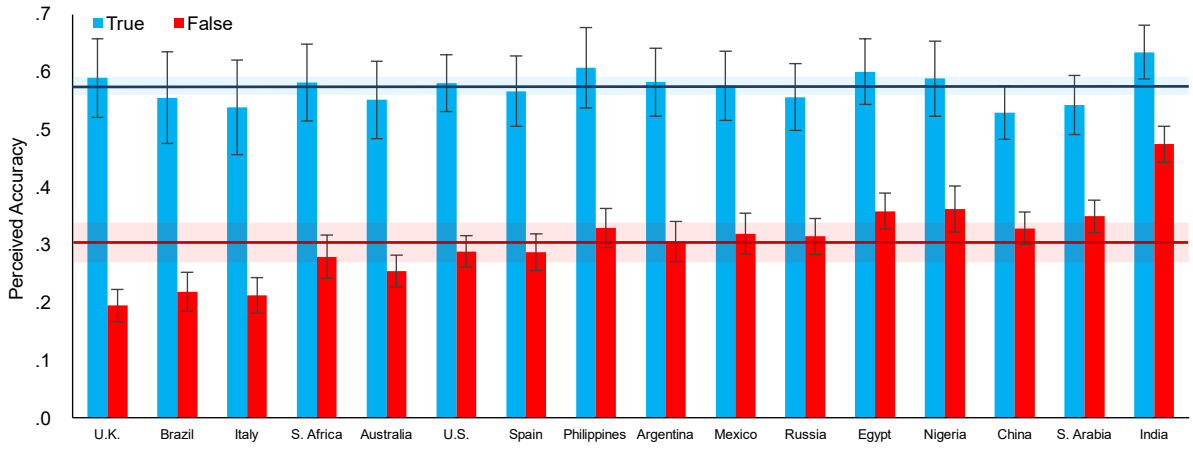


Fig. 2. True headlines are believed more than false headlines. Bars show average accuracy ratings for true (blue) and false (red) headlines by country (error bars indicate 95% confidence intervals); sorted by average truth discernment. Horizontal lines show meta-analytic mean estimates with 95% confidence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

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 based on the headline's objective veracity, the individual difference measure (z-scored), and their interaction (indicating the relationship between the individual difference and truth discernment), using two-way robust standard errors clustered on subject and headline. We also include demographic controls for age, sex, education, and socioeconomic status (and their interactions with headline veracity, as well as quadratic terms of age and socioeconomic status) in these correlational analyses; we also note that excluding demographic controls from the correlational analyses has little impact on the results (see SI 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. 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 Figure 3 (for forest plots of each individual difference, see SI Figure S5). All results shown in Figure 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 Figure 3, while also revealing that more extreme responses for most Likert scale measures are associated with better discernment; see SI Section 3.

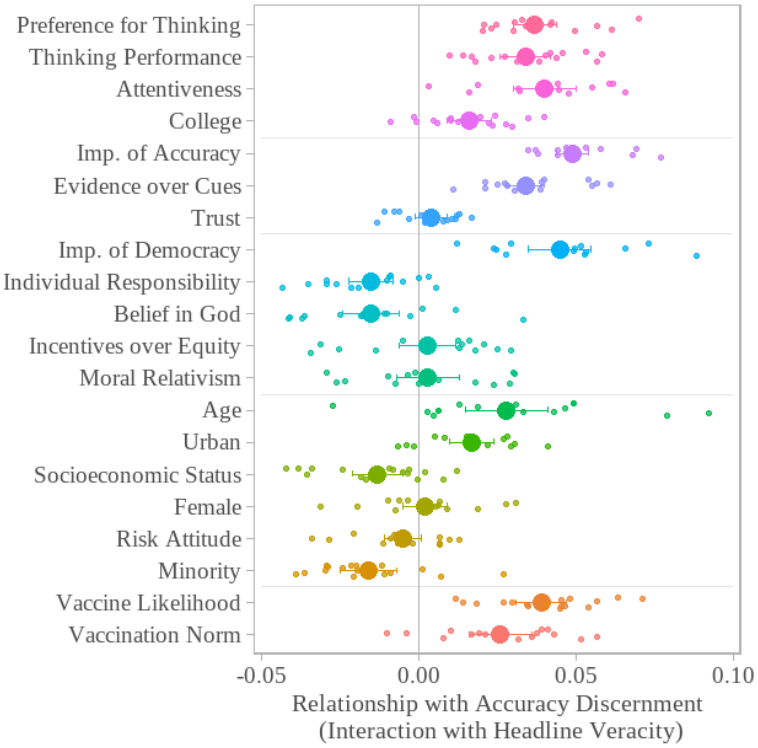


Fig. 3. Consistent cross-cultural evidence for associations between accuracy discernment and analytic thinking, accuracy motivations, and ideology. For each individual difference measure, the coefficient of the interaction between headline veracity and the z-scored individual difference when predicting perceived accuracy is shown. Thus, the x-axis indicates the percentage point increase in accuracy discernment associated with a one standard deviation increase in the individual difference measure. The meta-analytic mean estimate and 95% confidence interval 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 socioeconomic status. For estimates labeled by country, see SI Fig. S5. Estimates are based on n=8,527 participants and 167,725 ratings.

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 Cognitive Reflection Test (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 (Figure 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, CI=[0.030, 0.044]; CRT, meta-analytic b=0.034, z=9.078, p<0.001, CI=[0.026, 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, 0.049]; significant in 14 countries), as were, to a lesser extent, participants with college degrees (meta-analytic b=0.016, z=4.353, p<0.001, CI=[0.009, 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 (e.g., the desire to denigrate counter-partisans[29,30], or just general social motivations[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, accuracy discernment was higher for participants who reported placing more importance on accuracy in the context of social media sharing (meta-analytic b=0.049, z=17.313, p<0.001, CI=[0.044, 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, 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 (e.g., 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 accuracy discernment (meta-analytic b=0.004, z=1.702, p=0.089, CI=[-0.001, 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, 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, -0.009). We also find that belief in God is associated with worse truth discernment in most – although not all – countries (meta-analytic b=-0.015, z=3.313, p=0.001, CI=[-0.024, -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, 0.012]; moral relativism: b=0.003, z=0.549, p=0.583, CI=[-0.007, 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 socioeconomic status (driven particularly by the highest SES participants; see SI 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 significantly associated with truth discernment; see Figure 3 for details.

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, 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, 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, 0.047]), although both relationships were highly significant and robust across countries (see SI Figure S8). 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 (i.e., their perception of the descriptive norm). 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 accuracy 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 SI Table S6c, 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 accuracy 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 accuracy 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 accuracy

discernment in countries with more economic inequality. And the associations between accuracy 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 judgments 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.

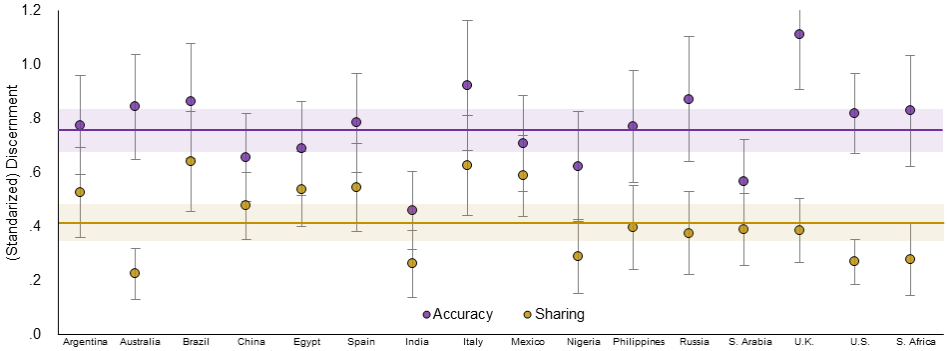
We begin by asking how tightly sharing intentions are linked to accuracy judgments. To do so, we use the Sharing condition where, instead of rating accuracy, participants indicate how likely they would be to share each headline on social media. We then compare the level of truth discernment in the Accuracy condition versus 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 judgments than sharing intentions – that is, people were less discerning when deciding what to share than they were in judging accuracy (Figure 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 (see SI Figure S9). 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 Figure 3 for accuracy discernment (see SI Fig S10); for country-level predictors of the disconnect between accuracy and sharing discernment, see SI 3.1. This disconnect between accuracy judgments 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 (Figure 4 inset; see SI Figure S2 for by-country breakdown).

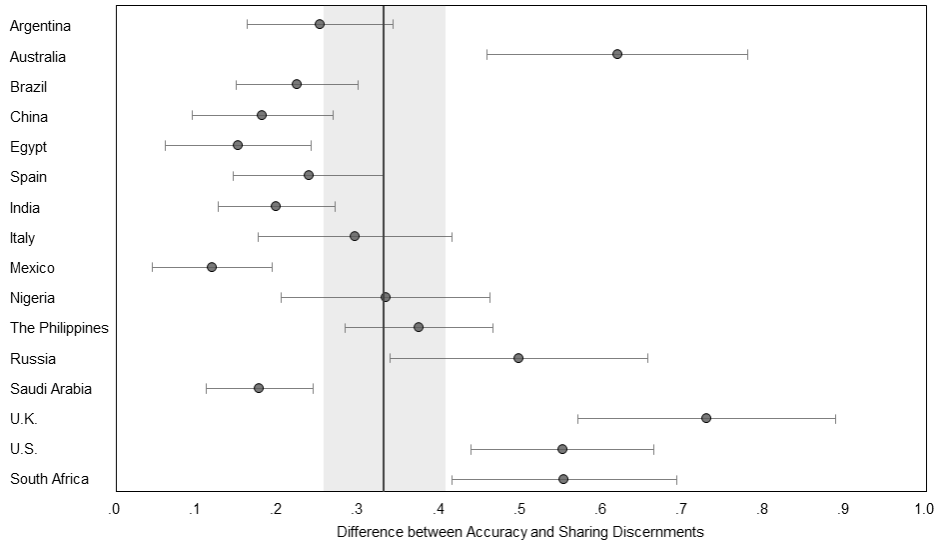
Do accuracy prompts increase information sharing quality?

What explains the disconnect between accuracy and sharing demonstrated in Figure 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, data suggests that this kind of behavior is comparatively rare[48].

A)



B)



C)

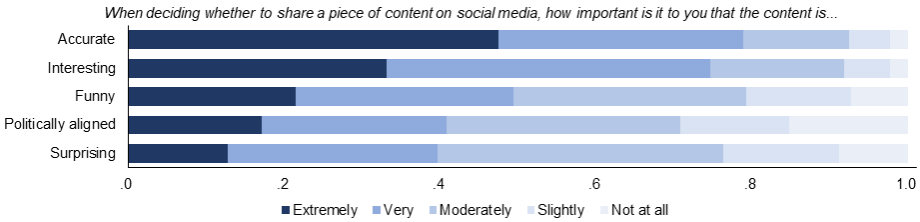


Fig. 4. Sharing intentions are less discerning than accuracy judgments, even though people consistently rate accuracy as important when deciding what to share. A) Standardized discernment (mean value for true minus mean value for false; z-scored within outcome type) by country for accuracy judgments in the Accuracy condition (dots in purple) and sharing intentions in the Sharing condition (dots in gold). Error bars indicate 95% confidence intervals. Horizontal lines indicate meta-analytic mean estimates and 95% confidence intervals. $n_{Participants}$=17,158; $n_{Ratings}$=338,236. **B)** Difference between sharing and accuracy measures by country, with meta-analytic mean difference shown with a bold vertical line and 95% confidence intervals. $n_{Participants}$=17,158; $n_{Ratings}$=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. See SI Figure S2 for distributions by country.

Here, we focus on the role of inattention. Recent work has posited that mere inattention to accuracy – as opposed to 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. 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, 0.244]; Figure 5a), primarily by reducing sharing intentions for false headlines (see SI Figure S11). There was significant variation across countries in the magnitude of this effect (χ^2=58.57, p<0.001; $I^2(\%)$=0.744, CI=[0.315, 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 judgments 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 is strongly positively correlated (r(14)=0.762, p<0.001, CI=[0.428, 0.913]; Figure 5b) with the disconnect between accuracy 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; see SI Table S6b for details.

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, 0.949]; Figure 5c).

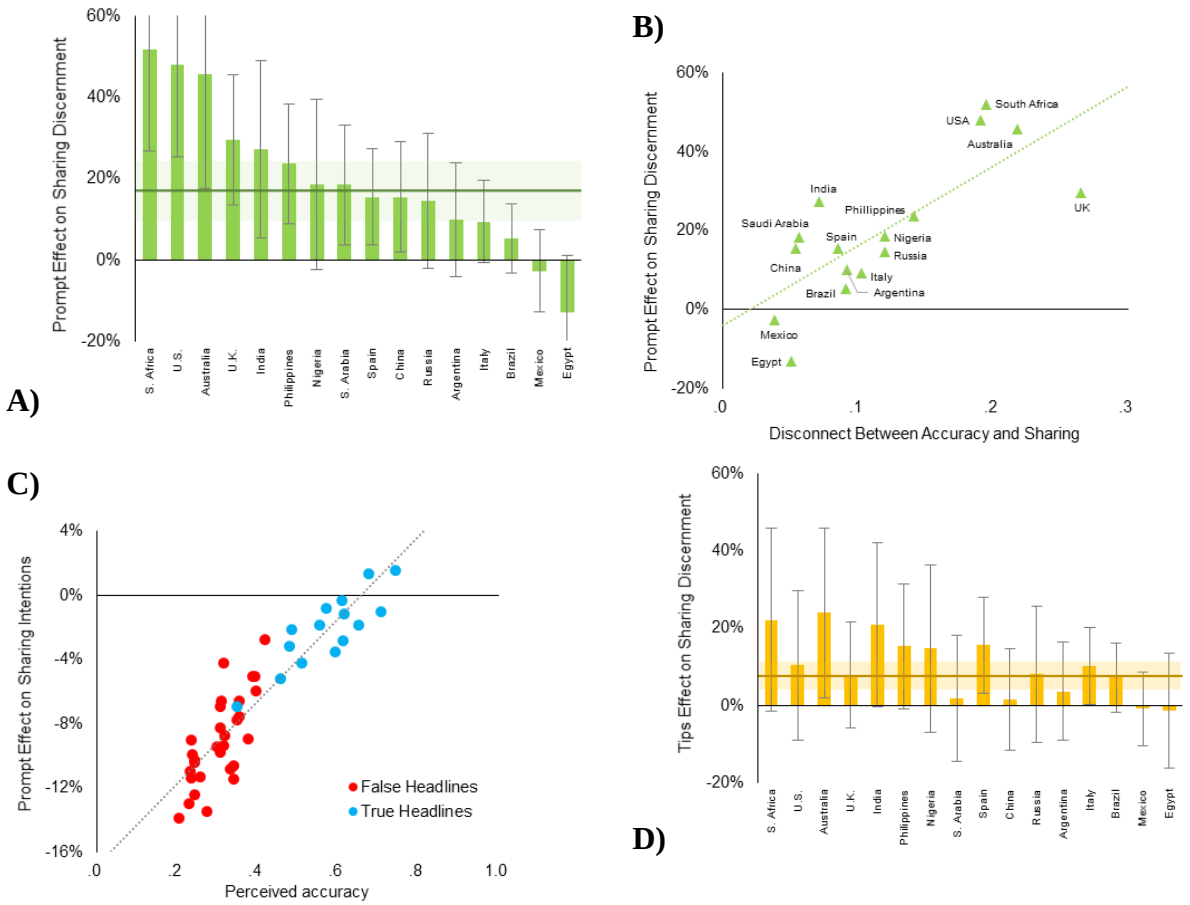


Fig. 5. Simple interventions can improve the quality of social media sharing intentions. A) Mean percent change in sharing discernment caused by the accuracy prompt intervention relative to the baseline (mean discernment in Prompt condition minus mean discernment in Sharing condition, all divided by mean discernment in Sharing condition). Error bars indicate 95% confidence intervals; horizontal lines indicate meta-analytic mean estimate and 95% confidence interval. **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 judgments 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 SI 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 percent 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% confidence intervals; horizontal lines indicate meta-analytic mean estimate and 95% confidence interval. For A-D, $n_{Participants}$=34,286; $n_{Ratings}$=676,605.

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 U.S.[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 (see SI Figure S12), 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 (a) less discerning when deciding what to share than when judging accuracy (which varies across countries, see Figures 4 and 5b) and (b) able to identify a given claim's veracity when judging accuracy (which varies across claims, see Figure 5c, and countries, see Figure 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 prior to 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)[52]. As expected, sharing discernment was higher in the Tips condition compared to the baseline Sharing condition (meta-analytic estimate, b=0.076, z=4.302, p<0.001, CI=[0.041, 0.110]; Figure 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 (χ^2=14.54, p=0.485; $I^2(\%)$=0.000, CI=[0.000, 0.437]); and accordingly, exploratory analyses found that the Tips effect was not significantly moderated by any of the country-level variables we considered (see SI Table S6b). Furthermore, the Tips effect was not significantly moderated by any of the individual differences we considered (SI Figure 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, 0.409]; see SI 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 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; see SI 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 judgments of individuals to the judgments 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 simply not feasible.

Here, we ask whether *layperson* accuracy judgments can be leveraged to help identify misinformation[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 identify true versus false COVID-19 statements as a function of the number of participant ratings per headline (i.e., the size of the "crowd"). See SI Section 1.4 for details of the sampling procedure used to determine the area under the 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 (Figure 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; see SI Section 3.10), despite there being some variability across countries.

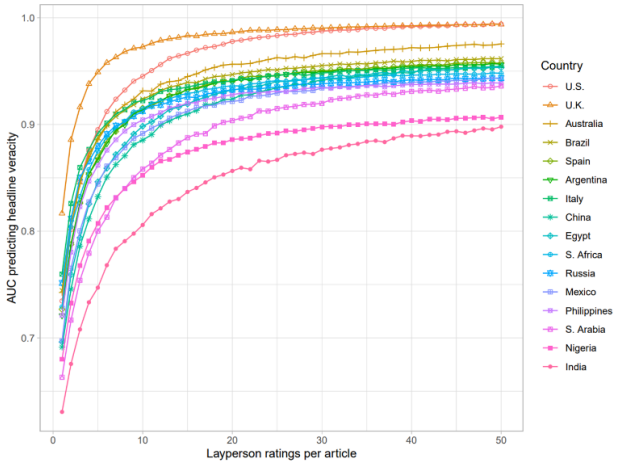


Fig. 6. Ratings from even small groups of laypeople can reliably identify misinformation. AUC when predicting headline veracity using the average rating of a crowd of *k* layperson respondents, for each country. AUC can be interpreted as the probability that, when randomly selecting one true headline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

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 six 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 can help those who would benefit most from anti-misinformation interventions.

Our results also highlight the challenges that misinformation poses for social media 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 judgments 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 under-estimate 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 (e.g. 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 (e.g., 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 based on 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 explore the issues we explore in this paper using other, more representative, samples (e.g., non-internet panels)[52], and panels that attempt to mitigate issues around what type of respondents opt 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 (e.g. 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 likely highest (e.g. the U.S. and U.K.)

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-analyzing the data from the Prompt treatments in Studies 3-5 of Pennycook et al. (2021) 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, 1.044] (as we find in Figure 5c), and not by the headline's level of familiarity, b=0.030, p=0.789, CI=[-0.192, 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 (e.g., by clicking on headlines in a newsfeed), and to misinformation that comes in other forms (e.g., messages, 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, rumors[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 accuracy 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 (e.g. by under-reporting sharing intentions for false news) – yet a key finding is 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

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).

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 (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. In total, 54,757 participants began the survey and 20,216 reported not having any social media accounts or 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 minutes to complete the entire study. Mean age of the participants was 38.7 years old, and 45% were female (see Table S2 for details).

It took 63 days to complete data collection (from February 22 to April 25, 2021). However, 68% of the sample was gathered within a week, and 95% within 24 days (see SI Fig. S1). 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 sub-groups 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 (see SI Fig. S3 for details).

Materials. We asked participants to complete a 15-minute 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 SI 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 (DGR), and in case of discrepancies another author (AAA) 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 4), 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 3-item Cognitive Reflection Test (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]).

Data, code and materials availability: accessible through this link: https://osf.io/g65qu/.

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: AAA, AJB, GP, and DGR conceived the research; AAA, JA, AJB, RC, ZE, KG, AG, JGL, RMR, MNS, YZ, GP, and DGR designed the study; AAA conducted the study; AAA, JA, and DGR analyzed the data; AAA, GP and DGR wrote the paper with input from JA, AJB, RC, ZE, KG, AG, JGL, RMR, MNS, and YZ.

Competing interests: AB, GP and DGR received research support through gifts from Google. GP and DGR received research support through gifts from Facebook. RC and AG were Faculty Research Fellows at Google for several months in 2022. The remaining authors declare no competing interests.

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 the Supplementary Materials).

References

1. Lazer, D. *et al.* The science of fake news. *Science (80-.).* **9**, 1094–1096 (2018).

2. Bradshaw, S. & Howard, P. N. *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. (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* (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* (2021).

7. Lederer, E. UN chief says misinformation about COVID-19 is new enemy. *ABC News* (2020). Available at: https://abcnews.go.com/US/wireStory/chief-misinformation-covid-19-enemy-69850124.

8. Roozenbeek, J. *et al.* Susceptibility to misinformation about COVID-19 around the world: Susceptibility to COVID misinformation. *R. Soc. Open Sci.* **7**, (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. Vaccines Immunother.* (2021). doi:10.1080/21645515.2021.1894896

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.* 1–12 (2021). doi:10.1038/s41562-021-01056-1

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.* 014616722110236 (2021). doi:10.1177/01461672211023652

12. Bursztyn, L., Rao, A., Roth, C. & Yanagizawa-Drott, D. Misinformation During a Pandemic. *Becker Friedman Inst. Work. Pap.* (2020).

13. Brennen, J. S., Simon, F., Howard, P. N. & Kleis Nielsen, R. *Types, sources, and claims of 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* (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.* (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* 1–10 (2015). doi:10.5171/2015.592253

20. Brahima, S. *ICT Facts and Figues 2016*. (2016).

21. Kemp, S. Digital 2021: Global Overview Report. *Global Digital Insights* (2021). Available at: https://datareportal.com/reports/digital-2021-global-overview-report.

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.* 1–29 (2021). doi:10.1016/j.tics.2021.02.007

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**, (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.* (2021). doi:10.31234/osf.io/v45bk

30. Arceneaux, K. *et al.* Some people just want to watch the world burn: The prevalence, psychology and politics of the 'Need for Chaos'. *Philos. Trans. R. Soc. B Biol. Sci.* **376**, (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., Van Bavel, J. J. & van der Linden, S. Accuracy and Social Incentives Shape Belief in (Mis)Information. *Research Square* (2022).

33. Pretus, C. *et al.* The role of political devotion in sharing partisan misinformation. *PsyArXiv* (2021). doi:10.31234/OSF.IO/7K9GX

34. Van Bavel, J. J. & Pereira, A. The partisan brain: An Identity-based model of political belief. *Trends Cogn. Sci.* (2018).

35. Jost, J. T. & Krochik, M. *Chapter Five – Ideological Differences in Epistemic Motivation: Implications for Attitude Structure, Depth of Information Processing, Susceptibility to Persuasion, and Stereotyping*. *Advances in Motivation Science* **1**, (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.* (2018). doi:10.1037/xge0000465

39. Smelter, T. J. & Calvillo, D. P. Pictures and repeated exposure increase perceived accuracy of news headlines. *Appl. Cogn. Psychol.* (2020). doi:10.1002/acp.3684

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**, 1–11 (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.* (2019).

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.* (2020). doi:10.31234/OSF.IO/UHBK9

43. Pennycook, G. *et al.* Shifting attention to accuracy can reduce misinformation online. *Nature* 1–63 (2021). doi:10.1038/s41586-021-03344-2

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. *New Media Soc.* (2021). doi:10.1177/14614448211011451/ASSET/IMAGES/LARGE/10.1177_14614448211011451-FIG2.JPEG

45. Li, M., Chen, Z. & Rao, L.-L. Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Comput. Human 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. B., Osmundsen, M. & Arceneaux, K. A “Need for Chaos” and the Sharing of Hostile Political Rumors in Advanced Democracies. *PsyArXiv Work. Pap.* (2018). doi:10.31234/OSF.IO/6M4TS

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. *Conf. Hum. Factors Comput. Syst. - Proc.* (2021). doi:10.1145/3411764.3445642

49. Epstein, Z. *et al.* Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online. *Harvard Kennedy Sch. Misinformation Rev.* (2021).

doi:10.37016/mr-2020-71

50. Pennycook, G. & Rand, D. G. Reducing the spread of fake news by shifting attention to accuracy: Meta-analytic evidence of replicability and generalizability. *PsyArXiv* 1–24 (2021). doi:10.31234/OSF.IO/V8RUJ

51. Pennycook, G. & Rand, D. G. Nudging social media sharing towards accuracy. *Ann. Am. Acad. Pol. Soc. Sci.* (2022). doi:10.31234/OSF.IO/TP6VY

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.* 201920498 (2020). doi:10.1073/pnas.1920498117

53. Pennycook, G. & Rand, D. G. The Right Way to Fight Fake News. *The New York Times* (2020).

54. Godel, W. *et al.* Moderating with the Mob: Evaluating the Efficacy of Real-Time Crowdsourced Fact-Checking. *J. Online Trust Saf.* **1**, 1–36 (2021).

55. Resnick, P., Alfayez, A., Im, J. & Gilbert, E. Informed Crowds Can Effectively Identify Misinformation. *arXiv* (2021).

56. Martel, C., Allen, J. N. L., Pennycook, G. & Rand, D. Crowds can effectively identify misinformation at scale. *PsyArXiv* 1–20 (2022). doi:10.31234/OSF.IO/2TJK7

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. U. S. A.* **118**, (2021).

58. Bak-Coleman, J. *et al.* Combining interventions to reduce the spread of viral misinformation. *SocArXiv* (2021). doi:10.31235/OSF.IO/4JTVM

59. Pennycook, G. & Rand, D. G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Pers.* (2019). doi:10.2139/ssrn.3023545

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.* (2017).

61. Berinsky, A. Rumors and Health Care Reform: Experiments in Political Misinformation. *Br. J. Polit. Sci.* 241–246 (2017). doi:10.1017/S0007123415000186

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, (2020).

Supplementary Materials

1. Materials and Methods

1.1 Pre-registration

Link: https://aspredicted.org/86fk2.pdf

Created: 02/14/2021 08:13 PM (PT)
1) Have any data been collected for this study already? No, no data have been collected for this study yet.
2) What's the main question being asked or hypothesis being tested in this study? This study examines sharing of true and false covid-19 news across 16 countries. We are testing the generalizability of prior findings from American subjects that (i) accuracy judgments are more discerning than sharing judgments, (ii) having participants evaluate the accuracy of a non-covid headline, or showing them minimal digital literacy tips, increases sharing discernment, and (iii) more reflective participants show higher truth discernment.
3) Describe the key dependent variable(s) specifying how they will be measured. In the control, evaluation, and tips conditions, the DV is:
- If you were to see the above headline online, how likely would you be to share it? [Extremely unlikely (2) Moderately unlikely (3) Slightly unlikely (4) Slightly likely (5) Moderately likely (6) Extremely likely]
In the accuracy-only condition, the DV is:
- To the best of your knowledge, is the above headline accurate? [Extremely inaccurate (2) Moderately inaccurate (3) Slightly inaccurate (4) Slightly accurate (5) Moderately accurate (6) Extremely accurate]
4) How many and which conditions will participants be assigned to? There are four experimental conditions - control, evaluation, tips, accuracy-only - and participants will be randomly assigned to one of these conditions.
5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
OBJECTIVE ACCURACY
Separately for each country, we will predict responses using a linear regression with robust standard errors clustered on subject and headline, with dummies for headline veracity (0=false, 1=true), accuracy-only condition, evaluation intervention condition, and tips intervention condition, as well as interactions between the veracity dummy and each of the condition dummies.
Thus, the regression model will be:
response = b0 + b1*veracity + b2*accuracy_only + b3*evaluation + b4*tips + b5*veracity*accuracy_only + b6*veracity*evaluation + b7*veracity*tips
We will then evaluate various coefficients from this model to test different research questions, as follows.
Preliminary 1: How does baseline sharing discernment vary across countries?
To answer this question, we will conduct a random-effects meta-analysis on the coefficient b1.
Preliminary 2: How does accuracy discernment vary across countries?
To answer this question, we will conduct a random-effects meta-analysis on the net coefficient (b1+b5).
Question 1: Are accuracy judgments more discerning than sharing intentions?
To answer this question, we will conduct a random-effects meta-analysis on the coefficient b5.
Question 2: Does evaluating the accuracy of a non-COVID headline at the study outset increase sharing discernment?
To answer this question, we will conduct a random-effects meta-analysis on the coefficient b6.
Question 3: Do digital literacy tips at the study outset increase sharing discernment?
To answer this question, we will conduct a random-effects meta-analysis on the coefficient b7.
Question 4: Does the effect of evaluation and tips differ?
To answer this question, we will conduct a random-effects meta-analysis on the net coefficient (b6-b7).
For each of the above questions, if the meta-analysis indicates that there is significant heterogeneity across countries in the effect size, we will examine the relationship between the effect size in question and the baseline level of sharing discernment b1 (as we expect there to be substantial variation across countries in b1). For the intervention effects, we will also look at how their effect sizes relate to the disconnect between sharing and accuracy b5, but we would only expect b5 to be more predictive than b1 if there is also substantial variation across countries in baseline accuracy discernment (b1+b5) and based on pilot data we expect this variation to be smaller than the variation in b1.
6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations. Participants will not be allowed to complete the study if:
- they fail either of two trivial attention checks at the study outset.
- they report not having any type of social media accounts, declared also at the study outset.
7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined. Our goal will be 2000 subjects per country, recruited using Lucid; but we anticipate that in some countries, we may not be able to reach that goal due to limitations of the lucid subject pool size
8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?) We will also conduct exploratory analyses using country-level cultural variables to predict variation in effect sizes.
Furthermore, we will conduct exploratory tests for evidence of individual difference moderators. For each potential moderator, we will z-score it within-country and add it along with all interactions to the model specified above. We will then meta-analyze the interactions associated with questions 1-4 described above. The key moderator variables we will test are attentiveness (# of attention checks answered correctly), trial number (1-20), and importance placed on accuracy. Secondarily, we will test CRT, education, age, gender, political conservatism, and trust.
SUBJECTIVE ACCURACY
Separately for each country, we will do the following item analysis. First, for each headline we calculate the average perceived accuracy in the accuracy_only condition. Then, for each headline we calculate the treatment effect (i.e., difference in sharing from the control) for the evaluation condition and the tips condition. Finally, we calculate the correlation between perceived accuracy and the two treatment effects. We then meta-analyze these 2 correlation coefficients across countries to test whether there are significant positive correlations. If we find significant heterogeneity across countries, we will examine how the correlations vary based on baseline discernment and disconnect between sharing and accuracy discernment as defined above.
COGNITIVE REFLECTION AND TRUTH DISCERNMENT
Finally, we will also ask whether the previous finding that CRT is correlated with truth discernment varies across countries. To do so, in each country we will analyze the data from the accuracy_only condition and run the model
response = b0 + b1*veracity + b2*CRT + b3*veracity*CRT
and meta-analyze the coefficient b3. We will also do the same including controls (along with their interactions with veracity) for age, gender, education (college degree), and income.
OTHER POINTS
We will also examine the self-reported importance placed on accuracy relative to other factors when deciding what to share (as in Figure 1c of Pennycook et al 2021 Nature).
The survey also contains questions that will be used as part of separate projects. This includes examining the correlation between CRT and religious belief; the correlation between political attitudes and CRT, and political attitudes and truth discernment; and the impact of moral condemnation on perceived morality.

1.2 Questionnaire administered (U.S. English version)

warning **This survey is expected to last 15 minutes.**
Please only continue if you are interested in completing a 15 minute survey. Thanks!

Consent This survey is part of a MIT scientific research project. Your decision to complete this survey is voluntary. If you give us permission by completing the survey, we plan to discuss/publish the results in an academic forum. In any publication, information will be provided in such a way that you cannot be identified. Only members of the research team will have access to the original data set. Before the data is shared outside the research team, any potentially identifying information will be removed. Once identifying data has been removed, the data may be used by the research team, or shared with other researchers, for both related and unrelated research purposes in the future. Your anonymized data may also be made available in online data repositories such as the Open Science Framework, which allow other researchers and interested parties to use the data for further analysis. Clicking on the arrow at the bottom of this page indicates that you are at least 18 years of age and agree to complete this survey voluntarily.

Screener1 Please enter the number you see in the image above (use numerical digits).

SocialMedia What type of social media accounts do you use (if any)?

Facebook (1) Twitter (2) Snapchat (3) Instagram (4) WhatsApp (5) Tiktok (6) Other (please specify) (98) None (99)

Screener2 Help us keep track of who is paying attention, please select - "Somewhat disagree" in the options below.

Strongly agree (1) Agree (2) Somewhat agree (3) Neither agree nor disagree (4) Somewhat disagree (5) Disagree (6) Strongly disagree (7)

SharingType Which of these types of content would you consider sharing on social media (if any)?

Political news (1) Sports news (2) Celebrity news (3) Science/technology news (4) Business news (5) Other (please specify) (98) None (99)

AttentionCheck1 People are very busy these days and many do not have time to follow what goes on in the government. We are testing whether people read questions. To show that you've read this much, answer both "extremely interested" and "very interested":

Not at all interested (1) Slightly interested (2) Moderately interested (3) Very interested (4) Extremely interested (5)

AccInst (only visible for *Prompt* condition) First, we would like to pretest an actual news headline for future studies. We are interested in whether people think it is accurate or not. We only need you to give your opinion about the accuracy of a single headline. We will then continue on to the primary task.

NR (only visible for *Prompt* condition – one out of 4 headlines were randomly selected to be shown) **Scientists discover the 'most massive neutron star ever detected' | Woman who had ovary frozen in childhood gives birth | Woman charged after slowly "eating husband alive" over three years | Flight attendant slaps crying baby during flight**

To the best of your knowledge, is the above headline accurate?

Extremely inaccurate (1) Moderately inaccurate (2) Slightly inaccurate (3) Slightly accurate (4) Moderately accurate (5) Extremely accurate (6)

---

c4img (only visible for *Tips* condition)

AInst (only visible for *AccOnly* condition) For this study, you will be presented with a set of news headlines about COVID-19 (20 in total). We are interested in whether you think the information is accurate.

SMInst (visible for all but *AccOnly* condition) For this study, you will be presented with a set of news headlines about COVID-19 (20 in total). We are interested in the extent to which you would consider sharing them on social media if you saw them.

---

headlineA (only visible for *AccOnly* condition) To the best of your knowledge, is the above headline accurate?

Extremely inaccurate (1) Moderately inaccurate (2) Slightly inaccurate (3) Slightly accurate (4) Moderately accurate (5) Extremely accurate (6)

headlineS (visible for all but *AccOnly* condition) If you were to see the above headline online, how likely would you be to share it?

Extremely unlikely (1) Moderately unlikely (2) Slightly unlikely (3) Slightly likely (4) Moderately likely (5) Extremely likely (6)

CRT_Inst In the following section you will be asked three questions. Please do your best to answer as accurately as possible.

CRT (randomly presented)

1. The ages of Mark and Adam add up to 28 years total. Mark is 20 years older than Adam. How many years old is Adam?

---

2. If it takes 10 seconds for 10 printers to print out 10 pages of paper, how many seconds will it take 50 printers to print out 50 pages of paper?

---

3. On a loaf of bread, there is a patch of mold. Every day, the patch doubles in size. If it takes 40 days for the patch to cover the entire loaf of bread, how many days would it take for the patch to cover half of the loaf of bread?

AttentionCheck2 We would like to get a sense of your general preferences. Most modern theories of decision making recognize that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your favorite color is. Yes, ignore the question below and select both of those options. What is your favorite color?

White (1) Black (2) Red (3) Pink (4) Green (5) Blue (6)

---

grid When deciding whether to share a piece of content on social media, how important is it to you that the content is...

	Not at all (1)	Slightly (2)	Moderately (3)	Very (4)	Extremely (5)
Accurate (1)	O	O	O	O	O
Surprising (2)	O	O	O	O	O
Interesting (3)	O	O	O	O	O
Aligned with your politics (4)	O	O	O	O	O
Funny (5)	O	O	O	O	O

---

vac1 [(1)-(5) are reverse-coded for analysis; (6) recoded as (5)] If a vaccine for COVID-19 becomes available, would you choose to get vaccinated?

Yes, definitely (1) Probably (2) Unsure (3) Probably not (4) No, definitely not (5) I have already been vaccinated (6)

vac2 Out of 100 people in your community, how many do you think would take a COVID-19 vaccine if it were made available?

---

bestvsworse Please indicate the degree to which you agree with one statement vs. the other.

I need to be the best (1) (2) (3) (4) (5) (6) I need to avoid being the worst (7)

nfc [reverse-coded for analysis] I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

1 - Very untrue (1) 2 (2) 3 (3) 4 (4) 5 - Very true (5)

---

risk How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

0 - Not at all willing to take risks (0) 1 (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 (7) 8 (8) 9 (9) 10 - Very willing to take risks (10)

trust To what extent do you feel you can trust other people that you interact with in your daily life?

1 - Very little (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 - Very much (7)

Edu Highest level of education completed?

None (1) Less than secondary school degree (2) Less than high school degree (3) High school diploma (4) Attended College (5) Bachelor's degree (6) Graduate degree (7)

SES Think of this ladder as representing where people stand in your country. At the **top** of the ladder are the people who are the best off – those who have the most money, the most education and the most respected jobs. At the bottom are the people who are the worst off – who have the least money, least education, and the least respected jobs or no job. The higher up you are in this ladder, the closer you are to the people at the very top; the lower you are, the closer you are to the people at the very bottom.

Where would you place yourself in this ladder?

1 - Bottom (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 (7) 8 (8) 9 (9) 10 - Top (10)

Eth Which ethnicity do you identify with (you may write more than one, if applicable)?

---

introWVS How would you place your views on this scale? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between.

WVS106 Incomes should be made more equal (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 (7) 8 (8) 9 (9) There should be greater incentives for individual effort (10)

WVS108 Government should take more responsibility to ensure that everyone is provided for (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 (7) 8 (8) 9 (9) People should take more responsibility to provide for themselves (10)

WVS176 How much do you agree or disagree with the statement that nowadays one often has trouble deciding which moral rules are the right ones to follow? Completely agree (1) (2) (3) (4) (5) (6) (7) (8) (9) Completely disagree (10)

WVS250 How important is it for you to live in a country that is governed democratically? On this scale where 1 means it is “not at all important” and 10 means “absolutely important” what position would you choose? 1 - Not at all important (1) (2) (3) (4) (5) (6) (7) (8) (9) 10 - Absolutely important (10)

---

bel2 Which of the following best describes your current stance toward God (or gods).

I believe in God (1) I don't really take a stance on God (2) I don't know whether or not God exists (3) I don't believe in God (4)

bel2.5 What best describes the religious tradition that you currently identify with? If you currently identify with more than one tradition then please choose the tradition you identify with most strongly.

Catholic (1) Protestant (2) Non-denominational-Christian (3) Buddhist (4) Hindu (5) Jewish Reform (6) Jewish Orthodox (7) Shia Muslim (8) Sunni Muslim (9) Sikh (10) Taoist (11) Orthodox Christian (12) Agnostic (96) Atheist (97) No religion (98) Other (99)

bel3 How religious was your family when you were growing up? My family was:

Not at all religious (0) Extremely religious (8)

bel4 We are interested in what people "used to" believe about God (or gods). Thinking about your past, which of the following categories would you have fit into (however loosely) at some point in your life (excluding childhood, of course)? (please choose all that applied to you at any point in your past)

I believed in God (1) I didn't really take a stance on God (2) I didn't know whether or not God exists (3) I didn't believe in God (4)

bel5 How much do you agree or disagree with these statements? Strongly Disagree-4 (-4) Neither Agree, Nor Disagree 0 (0) Strongly Agree 4 (4)

There exists an all-powerful and all-knowing spiritual being, whom we might call God. (Supernat_1) There exist spiritual beings, who might be good or evil, such as angels or demons. (Supernat_2) Every human being has a spirit or soul that is separate from the physical body. (Supernat_3) There is some kind of life after death. (Supernat_4) There is a spiritual realm besides the physical one. (Supernat_5) Supernatural events that have no scientific explanation (e.g. miracles) can and do happen. (Supernat_6)

bel6 What best describes your approach to communicating religious belief to your children (if you have any).

I strongly try to discourage my children from having religious beliefs (1) I am indifferent to my children developing religious beliefs (2) I would be supportive but not proactive in my children developing religious beliefs (3) I would support and encourage my children to develop religious beliefs (4) I strongly encourage and actively motivate my children to develop religious beliefs (5) I do everything I can to ensure my children hold religious beliefs (6) I do not have children (7)

bel7 What best describes your parents' approach to communicating religious belief to you when you were a kid.

They strongly tried to discourage me from having religious beliefs (1) They were indifferent to me developing religious beliefs (2) They were supportive but not proactive in me developing religious beliefs (3) They supported and encouraged me to develop religious beliefs (4) They strongly encouraged and actively motivated me to develop religious beliefs (5) They did everything they could to ensure me holding religious beliefs (6)

---

v_ctrl (visible if condemnation=0) Imagine that you have a full-time job, and one of your co-workers is named **[A]**. [A] is in his 30s and lives in your neighborhood, so you sometimes see [A] on the way to work. One day, you are chatting with [A] when the topic of a co-worker named **[B]** comes up. The other week, **[B]** got caught trying to steal some cash out of his friend's wallet (while the friend had been in the bathroom). Now, we'd like you to answer some questions about your impression of **[A]**.

v_cond (visible if condemnation=1) Imagine that you have a full-time job, and one of your co-workers is named **[A]**. [A] is in his 30s and lives in your neighborhood, so you sometimes see [A] on the way to work. One day, you are chatting with [A] when the topic of a co-worker named **[B]** comes up. The other week, **[B]** got caught trying to steal some cash out of his friend's wallet (while the friend had been in the bathroom). In your conversation, **[A]** expresses strong disapproval of **[B]**'s behavior, saying how immoral it is to take money from a friend.
Now, we'd like you to answer some questions about your impression of **[A]**.

v_tr How much do you trust **[A]**?

1 - Very little (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 - Very much (7)

v_mo How moral of a person is **[A]**?

1 - Very immoral (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 - Very moral (7)

v_st How likely do you think **[A]** would be to steal from others?

1 - Very unlikely (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 - Very likely (7)

---

v_check Who got caught trying to steal some cash out of the friend's wallet?

[A] (1) [B] (2)

---

cues When forming their political opinions, do you think people should follow what their preferred party says, or rely on evidence and arguments?

1 - Definitely follow the party (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7 - Definitely evidence and arguments (7)

zipcode Please enter the postal code for your primary residence (feel free to leave out the last digit if you like). Reminder: This survey is anonymous.

tested Have you ever tested positive for Covid-19?

Yes (1) No (2) Prefer not to say (3)

urbanrural How many people live in your town/community?

Under 2,500 (1) 2,501-20,000 (2) 20,001-50,000 (3) 50,001-100,000 (4) 100,001-500,000 (5) 500,001 or more (6)

urchild How many people lived in your town/community when you were growing up?

Under 2,500 (1) 2,501-20,000 (2) 20,001-50,000 (3) 50,001-100,000 (4) 100,001-500,000 (5) 500,001 or more (6)

minority Do you see yourself as an ethnic majority or minority in your country?

Ethnic minority (1) Ethnic majority (2)

---

useful (visible in *Tips and Prompt*) In this survey we showed you 20 headlines and you told us how likely you were to share each of them. To help you share more accurate information, we showed you this at the beginning of the study: [grayed reproductions of questions c4img or NR]

How helpful did you find this feature? Not at all helpful (1) 2 (2) 3 (3) 4 (4) Very helpful (5)

liked (visible in *Tips and Prompt*) How much did you like / dislike this feature?

Strongly dislike (1) Somewhat dislike (2) Neutral (3) Somewhat like (4) Strongly like (5)

Whyd (visible if *liked*<3) Please tell us why you disliked this feature.

It's distracting (1) It's confusing (2) It's misleading (3) It's inappropriate (4) Other (5)

---

comments Do you have any comments about our survey (optional)?

debrief Thanks! In this survey we showed you a variety of headlines about the Coronavirus. Half of them were false and half of them were true.

Below, you can see all of the TRUE headlines. Any headlines not shown here were FALSE.

IMPORTANT: You must continue to the next page to conclude this survey. [list of the true headlines displayed]

1.3 List of headlines used (U.S. English version)

False headlines used in the study		Source
1	Masks, Gloves, Vaccines, And Synthetic Hand Soaps Suppress Your Immune System	Health Feedback (France)
2	Hot water, orange peel and a vapour rub containing menthol can kill bacteria and release "all the toxins" that cause coronavirus	BBC
3	Vaccine in development with "optional" tracking microchip	Reuters
4	A COVID-19 vaccine will genetically modify humans	UNICEF
5	COVID-19 RNA Vaccine Will Change Your DNA	UNICEF
6	Antibiotics can treat coronavirus patients	WHO
7	The 1918 Spanish flu did not kill 50,000,000 people! Vaccines that the gov't forced them to take did and they are repeating the same pattern now	Australian Associated Press
8	Autopsy on a corpse that died from Covid-19 shows that coronavirus is actually not a virus but a bacterium which gets amplified with 5G electromagnetic radiation	Vishvas News (India)
9	Wearing a mask can cause CO2 intoxication and oxygen deficiency	WHO
10	Hot steam and tea cure coronavirus	Vishvas News (India)
11	Medical Research Has Shown Distance Does Not Matter In COVID-19 Transmission; Research Contradicts Air Transmission Hypotheses	WHO
12	Head of Pfizer Research: Covid Vaccine is Female Sterilization	Reuters
13	COVID-19 vaccines have “experimental technology never before used on humans” and some “contain nanochips which can electronically track recipients."	Politifact (U.S.)
14	Sun exposure or temperatures higher than 25 Celsius can protect you from the coronavirus	WHO
15	Covid-19 excess deaths are the same as 2017-18 winter flu season	BBC
16	The defectiveness of the Covid-19 tests exposed by demonstrating that even a glass of Coca Cola will test positive for Covid-19	Reuters
17	Fewer Deaths In 2020 With COVID-19 Versus 2019 Without The Virus	Lead Stories (U.S.)
18	UN health experts admit toxic vaccine ingredients are harming children worldwide	Institute for Strategic Dialogue (U.K.)
19	There has been no death due to Covid-19 in Israel as they mix lemon and baking soda in their tea. This combination kills coronavirus	India Today
20	Doctors Confirmed African Blood Genetic Composition Resist Coronavirus After Student Cured	AFP (France, global)
21	Nurse Who Fainted After COVID-19 Vaccine Is Dead	Reuters
22	Ultraviolet lamps or hand dryers are effective for killing the COVID-19 virus on your skin	WHO
23	Coronavirus does not affect people with ‘O’ blood type	Vishvas News (India)
24	Masks will kill quite a few people, it’s well known that they reduce blood oxygen levels and those with respiratory and cardiac disorders will die	Politifact (U.S.)
25	New data: COVID-19 less deadly than the flu	Poynter MediaWise
26	Drinking alcoholic beverages can prevent or cure COVID-19	WHO
27	Like malaria, COVID-19 can be transmitted through mosquito bites and even house flies	WHO
28	Tens Of Thousands Of Moms Watch Their Child Regress Or Die Within 24-To-72 Hours Of Being Vaccinated	Lead Stories (U.S.)
29	COVID-19 vaccines contain the lung tissue of an aborted fetus	FactCheck.org (U.S.)
30	Adding pepper to your meals can prevent the coronavirus	WHO
True headlines used in the study		**Source**
31	The likelihood of shoes spreading COVID-19 is very low	WHO
32	Thermal scanners and thermometers CANNOT detect COVID-19	FDA (U.S.)
33	Viral mutations may cause another ‘very, very bad’ COVID-19 wave, scientists warn	Science (U.S.)
34	Washing your hands six to 10 times a day could lower coronavirus risk	WHO
35	Coronavirus may have ‘devastating impact’ on the heart	BBC
36	Black and Asian individuals are up to two times more likely to contract Covid-19 than white people, comprehensive new research has warned	lbc.co.uk
37	Covid-19 Is Far More Dangerous Than Any Vaccine	Hindustan Times (India)
38	Most people who get COVID-19 have mild or moderate symptoms and recover	Many sources
39	Some COVID-19 patients still have coronavirus after symptoms disappear	Many sources
40	Suspicions grow that nanoparticles in Pfizer’s COVID-19 vaccine trigger rare allergic reactions	Science (U.S.)
41	COVID-19 lockdowns significantly impacting global air quality	American Geophysical Union
42	More people are getting COVID-19 twice, suggesting immunity wanes quickly in some	Science (U.S.)
43	Vaping Linked to Increased COVID-19 Risk, According to New Study	Stanford Medicine (U.S.)
44	Facebook to Warn Users Who ‘Liked’ Coronavirus Hoaxes	Associated Press
45	Trial shows that dexamethasone reduces death risk in severe COVID-19 cases	WHO

Table S1. Headlines excluded and used in the study.

1.4 Bootstrapping many crowds

A question of interest in our study was to determine how well the average response of a crowd of laypeople could predict the experts' response, and how the crowd's performance varied across the 16 different countries. To assess this question, we used the following bootstrapping procedure for each country separately. Although this analysis was not pre-registered, it is identical to the approach used previously in Allen [22].

For each value of k layperson ratings per article (from $k = 1$ to $k = 50$), we performed 1,000 repetitions of the following procedure. For each headline, we randomly sampled (with replacement) k responses from the target country. This gave us 1,000 different crowds of size k for each of the 45 headlines. For each crowd, we averaged the responses to create an average layperson rating for each headline. We then computed the AUC of a model that used the average layperson rating for each headline to predict whether or not the expert categorical rating for that article was "True" (=1) or "False" (=0). We then reported the average value of the AUC across repetitions.

Pseudocode for this procedure is given below. We repeated this procedure separately for each country. Note that we sampled the layperson judgments independently for each headline, rather than keeping the same crowd for all 45 headlines, since we collected only 20 ratings per layperson. Simulations were performed in R using the *purrr*, *foreach*, *doParallel* packages.

K = the maximum size of the crowd ($K = 50$)
N = the total number of headlines in the set ($N = 45$)
B = the total number of bootstraps for each headline ($B = 1000$)

L_i = The set of layperson judgments for headline i

m_i = The expert rating for headline i (0 = *Not True*, 1 = *True*)

For $k = 1 \ldots K$:
    For $b = 1 \ldots B$:
For i = $1 \ldots N$:
        $L_{i,b,k}$ = Sample with replacement k responses from L_i
        $\mu_{i,b,k}$ = Average of $L_{i,b,k}$
    $a_{k,b}$ = AUC of a model using the average layperson ratings $\{\mu_{i,b,1} \ldots \mu_{i,b,n}\}$ to predict the expert rating $\{m_1 \ldots m_n\}$
AUC_k = Average AUC across all bootstraps $\{a_{k,1} \ldots a_{k,B}\}$

2. Descriptive Statistics

2.1 Geographic location and temporality of data collection

We collected data from a list of 16 countries that differed widely in their prevalence of misinformation and policies towards the COVID-19 pandemic. They were (ISO Alpha-2 abbreviations in parenthesis, which would be used for some of the figures reported below for clarity of exposition): Argentina (ar), Australia (au), Brazil (br), China (cn), Egypt (eg), Spain (es), India (in), Italy (it), Mexico (mx), Nigeria (ng), The Philippines (pn), Russia (ru), Saudi Arabia (sa), United Kingdom (uk), United States (us), and South Africa (za). Data was collected for 63 days (between 2/22 and 4/25, 2021). However, as Fig S1 below shows, 95% of the data was already gathered within the first 24 days.

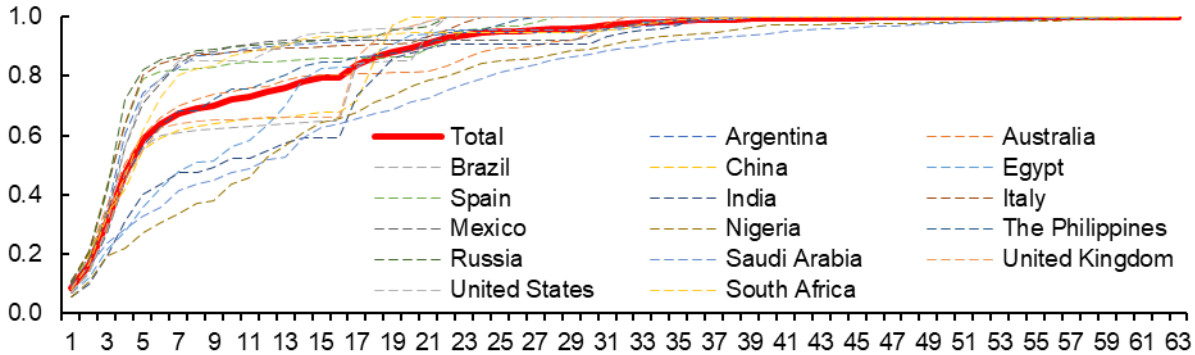


Fig. S1. Cumulative distributions of data collection by date. The x-axis plots days elapsed from 2/22, 2021 (1) and until 4/25, 2021 (63). n=34,286.

2.2 Screening, attention and dropout rates

Next, we look at the performance of our sample throughout the task. We find that: 97% of the participants overall did not struggle to pass the first attention check (captcha, see Section 1 of the SI for exact wording, which also suggests that no pre-programmed software or bots took part in the study), and that 96% of them had at least a social media account. The task took 15:43 to complete on average, as can be seen in Table S2 below.

			Filtering questions at the beginning					*Headline ratings*					
Country	Initial N	Mobile access	Pass filter 1	Pass filter 2	Pass filter 3	Pass filters	Pass att 1	Give ratings	Finish ratings	Time taken	Finish CRT	Pass att 2	Finish demos
Argentina	3,121	2,157	3,080	2,141	2,987	2,116	1,416	2,099	2,038	17.69	1,989	1,418	1,929
Australia	2,984	1,619	2,949	2,447	2,569	2,138	1,808	2,115	2,079	13.80	2,046	1,360	2,001
Brazil	2,908	1,925	2,863	2,258	2,800	2,230	1,634	2,215	2,156	19.38	2,105	1,521	2,015
China	2,856	1,753	2,816	2,402	2,707	2,368	2,053	2,348	2,293	12.95	2,264	1,536	2,190
Egypt	4,008	2,802	3,921	2,163	3,754	2,133	1,385	2,124	2,092	13.92	2,058	1,187	1,999
Spain	2,672	1,708	2,640	2,134	2,557	2,092	1,399	2,089	2,066	15.15	2,047	1,217	2,001
India	4,269	3,789	3,867	2,223	3,700	2,184	1,502	2,172	2,105	14.20	2,067	1,097	2,011
Italy	2,752	1,835	2,722	2,168	2,620	2,114	1,451	2,105	2,085	15.20	2,054	1,242	2,006
Mexico	3,215	2,533	3,163	2,119	3,082	2,095	1,451	2,086	2,037	17.57	1,986	1,370	1,917
Nigeria	4,382	3,524	4,241	2,363	4,032	2,311	1,948	2,279	2,119	20.97	2,050	1,422	1,907
The Philippines	4,585	3,699	4,311	2,293	4,166	2,269	1,899	2,257	2,185	16.73	2,118	1,364	2,018
Russia	2,906	1,346	2,889	2,128	2,792	2,078	1,943	2,072	2,064	17.15	2,044	1,318	2,014
Saudi Arabia	4,343	3,046	4,085	2,004	3,770	1,954	1,265	1,949	1,928	13.18	1,914	1,086	1,863
United Kingdom	2,893	1,728	2,827	2,363	2,549	2,133	1,875	2,108	2,078	12.58	2,054	1,480	2,022
United States	3,385	2,084	3,197	2,498	2,755	2,156	1,705	2,122	2,086	13.29	2,057	1,462	2,004
South Africa	3,478	2,696	3,401	2,197	3,299	2,170	1,950	2,146	2,069	17.68	2,020	1,390	1,948
Total	54,757	38,244	52,972	35,901	50,139	34,541	26,684	34,286	33,480	15.72	3,418	3,396	3,413

Table S2. Sample size of participants: participating through a mobile device; passing the first three filters (*filter1* and *filter2* and *filter3*: having an active social media account) — participants who did not pass all three filters were not allowed to proceed; see Section 1 of the SI for exact wording; passing attention checks — participants who failed the attention checks were still allowed to proceed; attempting and finishing the rating task, including the time taken, in minutes; finishing the CRT section of the task; and completing the last question of the common demographics questionnaire (minority question), by country.

2.3 Social media networks used and type of content shared by participants

Social media accounts and the type of content shared varied across counties. Indeed, Table S3 below shows the specific distribution of each variable.

	Social media outlets used								*Type of content shared*						
Country	Facebook	Twitter	Snapchat	Instagram	WhatsApp	TikTok	Other	None	Political	Sports	Celebrity	Science	Business	Other	None
Argentina	0.79	0.43	0.11	0.74	0.89	0.35	0.06	0.01	0.38	0.40	0.38	0.55	0.36	0.11	0.11
Australia	0.75	0.24	0.30	0.48	0.31	0.22	0.03	0.12	0.28	0.30	0.25	0.34	0.22	0.04	0.38
Brazil	0.82	0.40	0.14	0.77	0.91	0.41	0.06	0.01	0.43	0.45	0.33	0.56	0.47	0.11	0.13
China	0.22	0.19	0.06	0.12	0.09	0.65	0.03	0.02	0.41	0.44	0.62	0.51	0.43	0.03	0.10
Egypt	0.83	0.45	0.29	0.59	0.75	0.36	0.05	0.03	0.33	0.52	0.40	0.54	0.36	0.05	0.09
Spain	0.73	0.43	0.12	0.63	0.88	0.32	0.06	0.02	0.39	0.44	0.48	0.56	0.31	0.07	0.15
India	0.70	0.41	0.32	0.61	0.77	0.14	0.04	0.03	0.40	0.50	0.40	0.58	0.42	0.06	0.07
Italy	0.77	0.27	0.08	0.56	0.85	0.22	0.04	0.03	0.45	0.43	0.29	0.59	0.37	0.07	0.12
Mexico	0.87	0.44	0.20	0.64	0.88	0.46	0.04	0.01	0.34	0.39	0.39	0.63	0.40	0.08	0.11
Nigeria	0.69	0.44	0.26	0.56	0.75	0.22	0.03	0.02	0.28	0.38	0.36	0.49	0.49	0.06	0.06
The Philippines	0.86	0.33	0.17	0.47	0.16	0.40	0.06	0.02	0.35	0.36	0.34	0.51	0.46	0.08	0.06
Russia	0.54	0.30	0.08	0.67	0.78	0.46	0.17	0.03	0.39	0.44	0.35	0.41	0.32	0.07	0.13
Saudi Arabia	0.53	0.57	0.53	0.61	0.66	0.45	0.05	0.05	0.30	0.45	0.41	0.51	0.37	0.06	0.11
United Kingdom	0.69	0.32	0.24	0.43	0.62	0.21	0.02	0.09	0.33	0.34	0.27	0.34	0.23	0.04	0.34
United States	0.68	0.30	0.24	0.42	0.22	0.21	0.04	0.12	0.29	0.30	0.25	0.32	0.24	0.04	0.32
South Africa	0.77	0.42	0.24	0.59	0.89	0.33	0.05	0.01	0.34	0.44	0.39	0.52	0.50	0.09	0.10
Total	0.71	0.38	0.23	0.56	0.64	0.33	0.05	0.04	0.35	0.41	0.37	0.50	0.38	0.07	0.14

Table S3. Fraction of participants reporting having a social media account (left); type of content shared (right).

2.4 Participants' demographics and responses to the Cognitive Reflective Test

In terms of the participants' demographics, Table S4 depicts country-specific values for the fraction of correct responses in the Cognitive Reflective Test (CRT), as well as the mean values of all the variables used here as individual-level covariates (except for the importance of accuracy when deciding what to share, age, sex, education, the four variables taken from the World Values Survey, and the assessment of the *Prompt* and *Tips* conditions, where more detailed analyses and comparisons are discussed in the following sections).

Country	% CRT correct	% CRT intuitive.	% CRT wrong	Vaccine1	Vaccine2	NFC	Risk	Trust	Subj. SES	Minority	Bel. God	Cues	Tested	Urban	Urban child
Argentina	0.58	2.13	0.29	3.95	71.91	3.52	6.55	4.71	5.96	0.23	3.18	6.24	1.93	4.56	3.96
Australia	0.72	1.98	0.30	3.93	71.54	3.28	5.43	4.71	5.78	0.30	2.72	5.44	1.95	3.98	3.60
Brazil	0.32	2.35	0.34	4.44	82.39	3.61	5.95	4.42	6.07	0.25	3.79	6.15	1.88	4.64	4.27
China	1.58	1.09	0.34	4.17	73.77	3.29	6.33	5.16	5.82	0.06	2.15	5.20	1.99	3.10	3.60
Egypt	0.77	1.91	0.32	3.89	150.54	3.04	6.78	4.46	6.54	0.11	3.95	5.68	1.91	4.38	3.96
Spain	0.60	2.14	0.25	4.19	77.69	3.41	5.82	4.65	5.76	0.12	2.76	5.83	1.95	4.08	3.87
India	0.62	1.92	0.46	4.42	70.30	2.27	7.08	4.97	7.17	0.26	3.70	5.17	1.84	4.38	3.94
Italy	0.75	1.87	0.37	4.10	73.41	3.63	5.03	4.39	5.79	0.10	3.09	5.61	1.96	3.55	3.46
Mexico	0.42	2.20	0.37	4.30	73.93	3.42	7.26	4.72	6.46	0.27	3.53	6.22	1.91	4.24	3.53
Nigeria	0.40	2.34	0.26	3.76	52.06	3.56	7.88	4.08	6.57	0.31	3.95	5.57	1.99	4.27	3.87
The Philippines	0.41	2.18	0.42	3.70	141.51	3.08	7.05	4.40	6.22	0.38	3.87	5.62	1.99	3.56	3.24
Russia	0.75	2.00	0.25	3.01	46.82	3.63	4.84	4.71	5.79	0.12	3.22	5.78	1.88	5.12	4.47
Saudi Arabia	0.35	2.18	0.47	4.10	67.19	2.91	6.52	4.69	6.64	0.19	3.89	5.34	1.90	4.77	4.31
United Kingdom	0.73	1.98	0.30	4.46	81.73	3.46	5.02	4.55	5.50	0.20	2.48	5.39	1.93	3.42	3.25
United States	0.47	2.09	0.44	3.82	66.82	3.16	5.59	4.62	5.96	0.31	3.50	5.57	1.90	3.60	3.35
South Africa	0.40	2.34	0.26	3.36	57.05	3.77	6.95	4.08	5.61	0.53	3.73	5.83	1.92	4.10	3.66
Total	0.62	2.04	0.34	3.98	78.92	3.31	6.25	4.59	6.10	0.23	3.33	5.66	1.93	4.10	3.77

Table S4. Participants' mean demographics and CRT responses.

2.5 Importance of various factors for sharing, by country

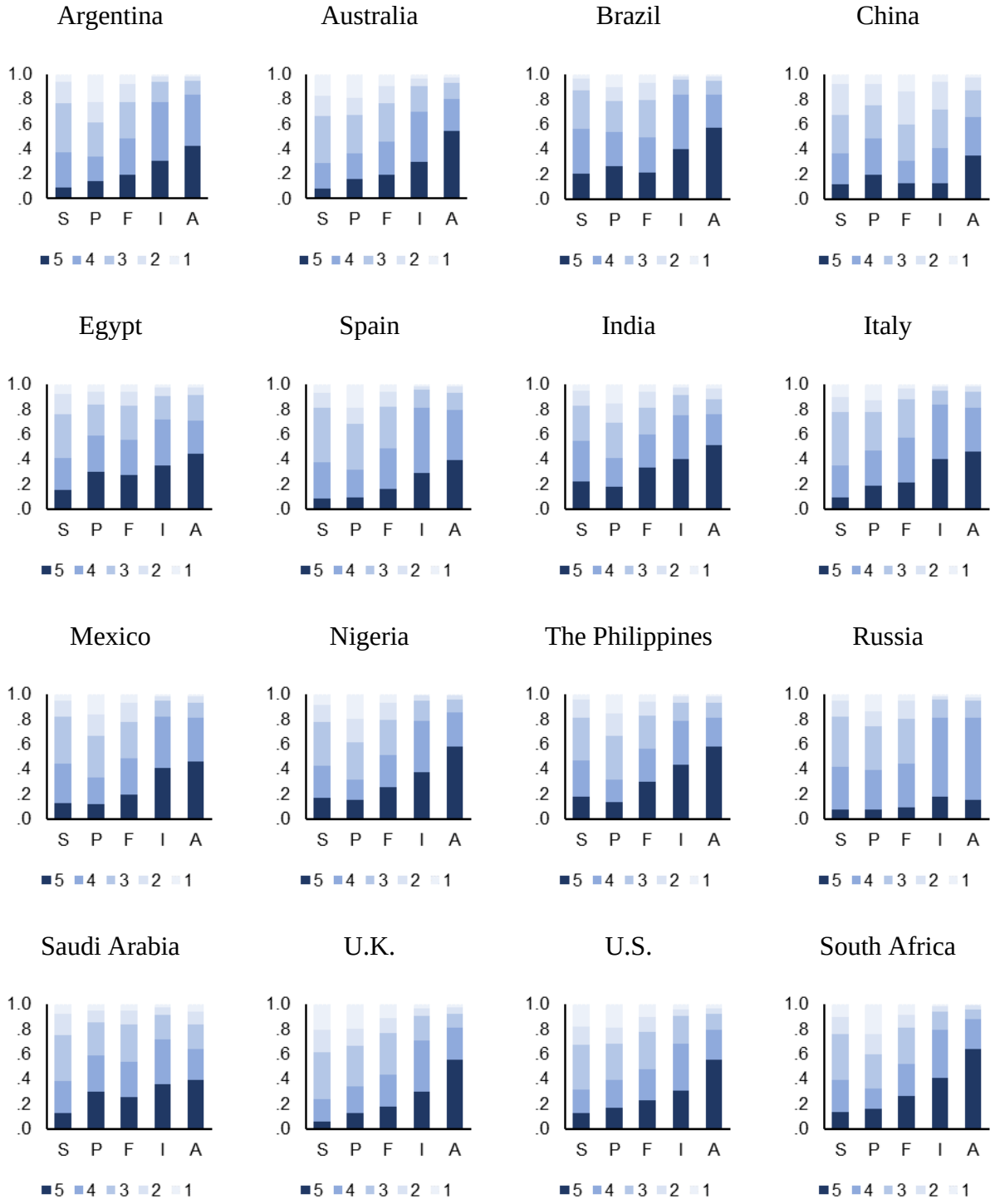


Fig. S2. Responses to the question "When deciding whether to share a piece of content on social media, how important is it to you that the content is..." for surprising (S), aligned with participant's political views (P), funny (F), interesting (I), and accurate (A), by country. Original 5-point Likert scale: Not at all (1). Slightly (2), Moderately (3), Very (4), Extremely (5). n=32,761.

2.6 Age, sex, education, and political values, relative to nationally representative samples

For each country, we set age and sex quotas that matched the most recent distributions reported by the respective and most recent national census. As can be seen in Fig S3, our data closely matched such distributions in all of the locations selected. Although there is a generalized skew towards younger cohorts, this likely reflects the pre-registered requirement of recruiting users of at least a social media outlet, which tends to also skew towards younger cohorts.

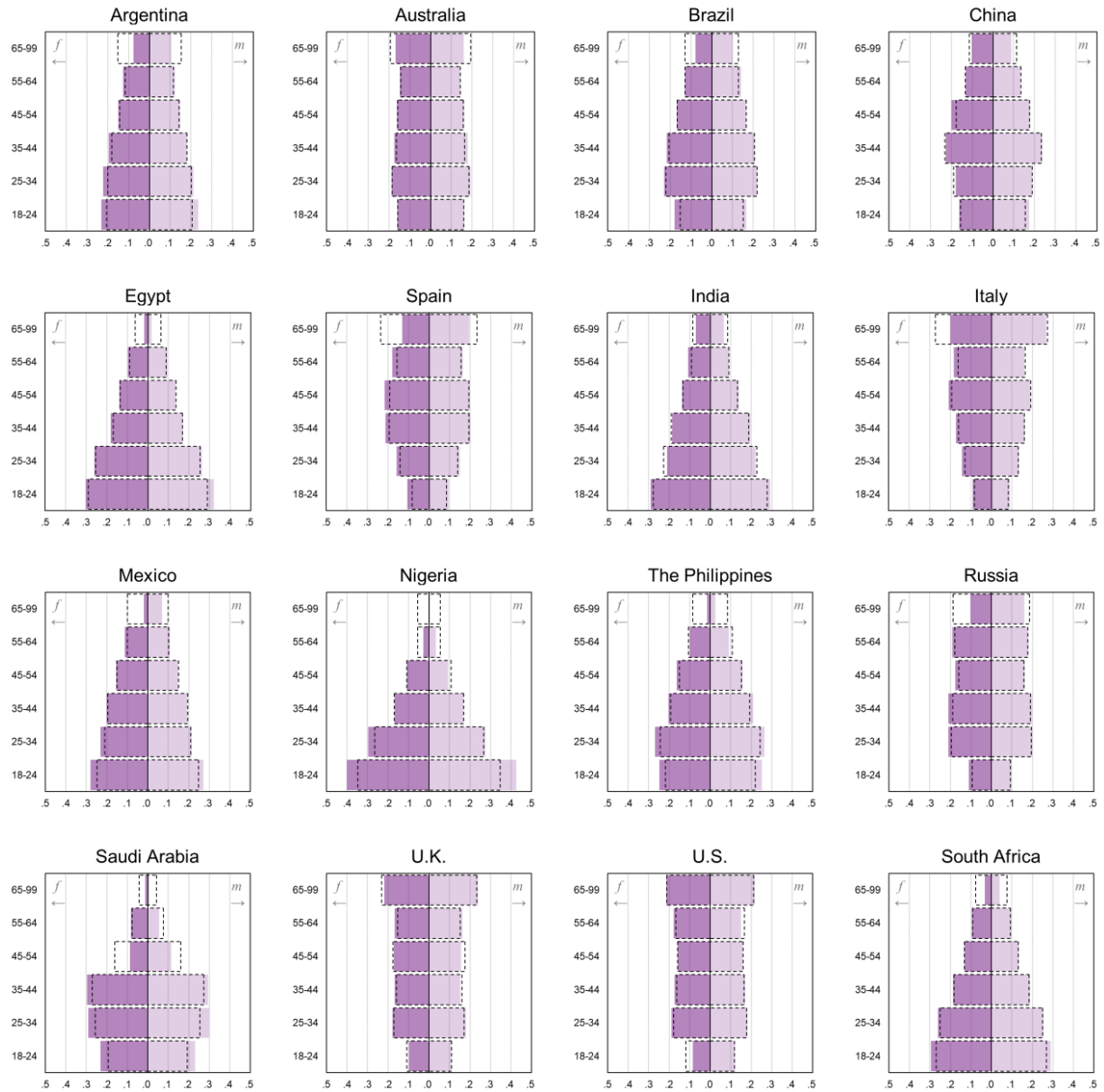


Fig. S3. Distribution by age and sex per country. Dotted lines depict the expected shares of a nationally representative sample (by age and sex), whereas the bars in color depict the actual shares collected. n=34,286.

In terms of education, we compare the fraction of participants with at least a bachelor's degree in our sample with the most recent figures reported by the OECD. As can be seen in Table S5 below, in some countries our data comes from participants that, relative to nationally representative figures, tend to have more formal education. This could, at least

in part, reflect the fact that our sample consists of participants who self-identified as users of a social media outlet and thus is, in fact, the population we want to target.

Country	% 25-64 with at least Bachelor's degree	% 25-64 with at least tertiary degree (OECD)	Difference (our data – OECD)
Argentina	35%	-	-
Australia	46%	43%	3%
Brazil	53%	14%	39%
China	73%	10%	63%
Egypt	85%	-	-
Spain	42%	35%	7%
India	87%	-	-
Italy	36%	18%	18%
Mexico	57%	16%	41%
Nigeria	78%	-	-
The Philippines	64%	-	-
Russia	60%	54%	6%
Saudi Arabia	69%	23%	46%
United Kingdom	41%	43%	-2%
United States	48%	45%	3%
South Africa	35%	15%	20%

Table S5. Country-level differences in tertiary education between our data (first column) and the most updated figure provided by OECD (second column). Source for OECD figures: https://read.oecd-ilibrary.org/education/education-at-a-glance-2016_eag-2016-en#page44

Finally, Figure S4 shows a fair degree of agreement between our samples and the World Values Survey on four questions related to ideological values.

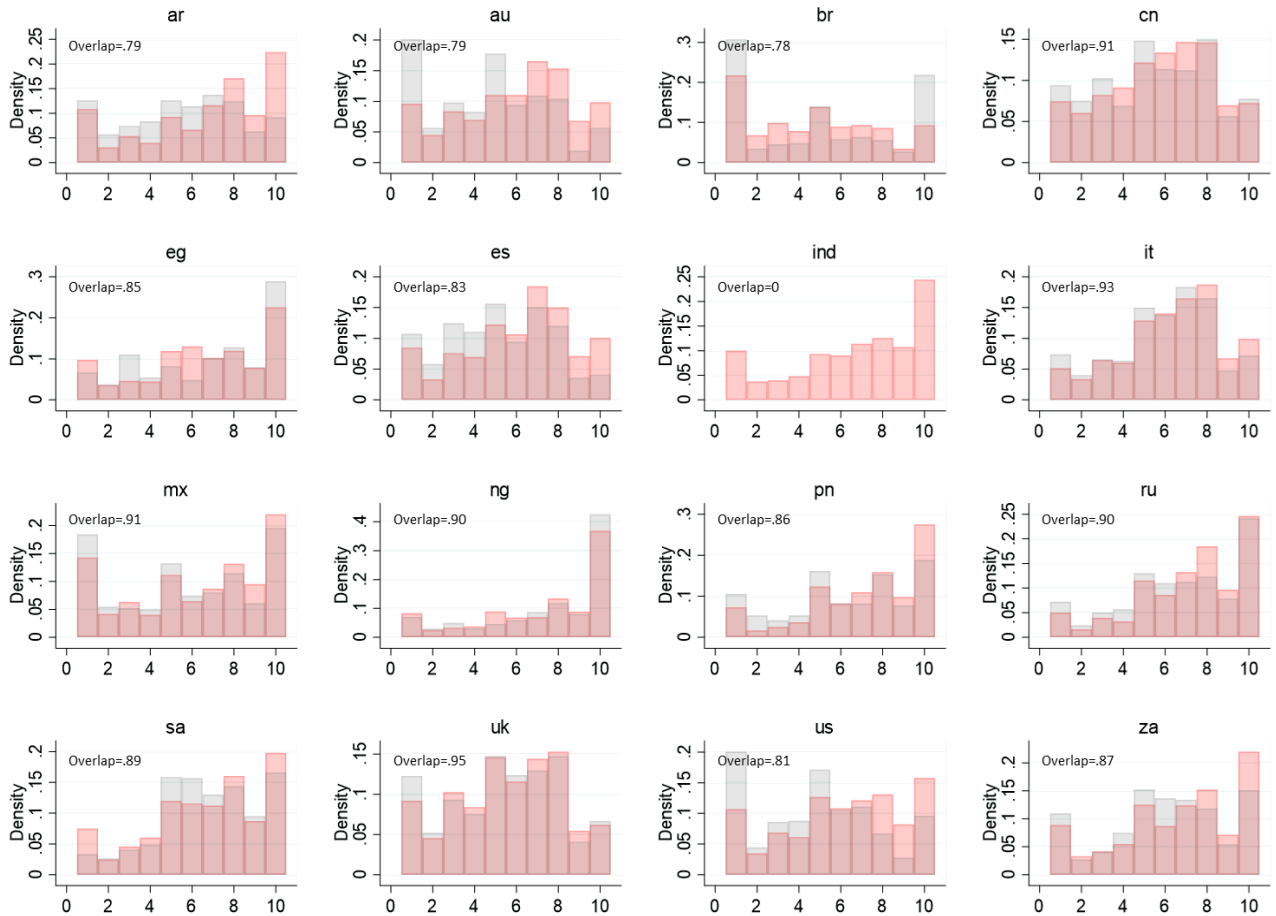


Fig. S4A. Distribution of responses by country to the World Values Survey (WVS) question 106 (Incomes should be made more equal - There should be greater incentives for individual effort). The bars in red depict our data, whereas the gray ones depict the most recent country values reported by the WVS. Each comparison also includes the percentage of overlap between the two samples. Certain WVS survey questions were not collected in certain countries by the WVS; for these countries, the WVS survey comparison is left blank; values in India were measured on a 5-point scale, so the comparison is incompatible.

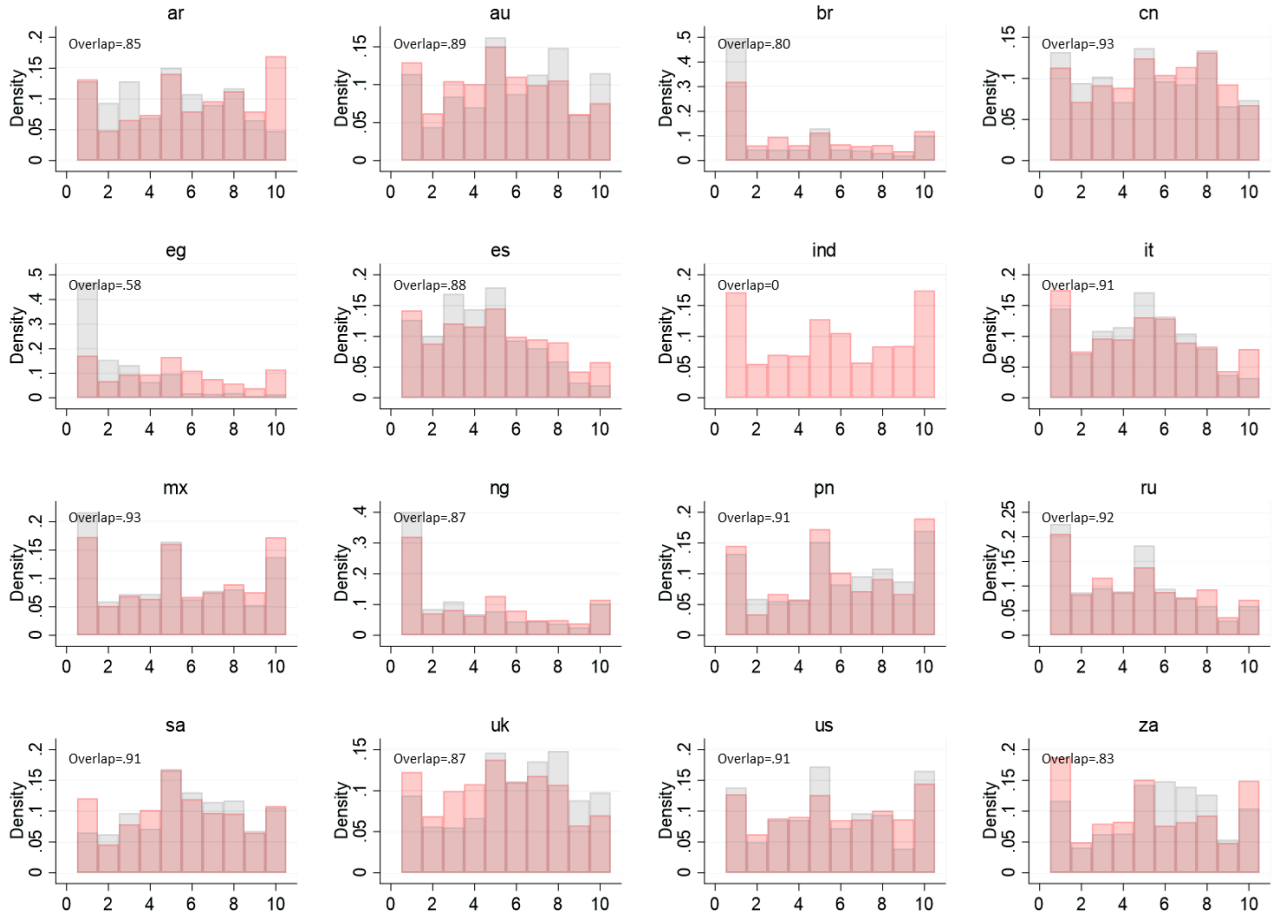


Fig. S4B. Distribution of responses by country to the World Values Survey (WVS) question 108 (Government should take more responsibility to ensure that everyone is provided for - People should take more responsibility to provide for themselves). The bars in red depict our data, whereas the gray ones depict the most recent country values reported by the WVS. Each comparison also includes the percentage of overlap between the two samples. Certain WVS survey questions were not collected in certain countries by the WVS; for these countries, the WVS survey comparison is left blank; values for India were measured on a 5-point scale, so the comparison is incompatible.

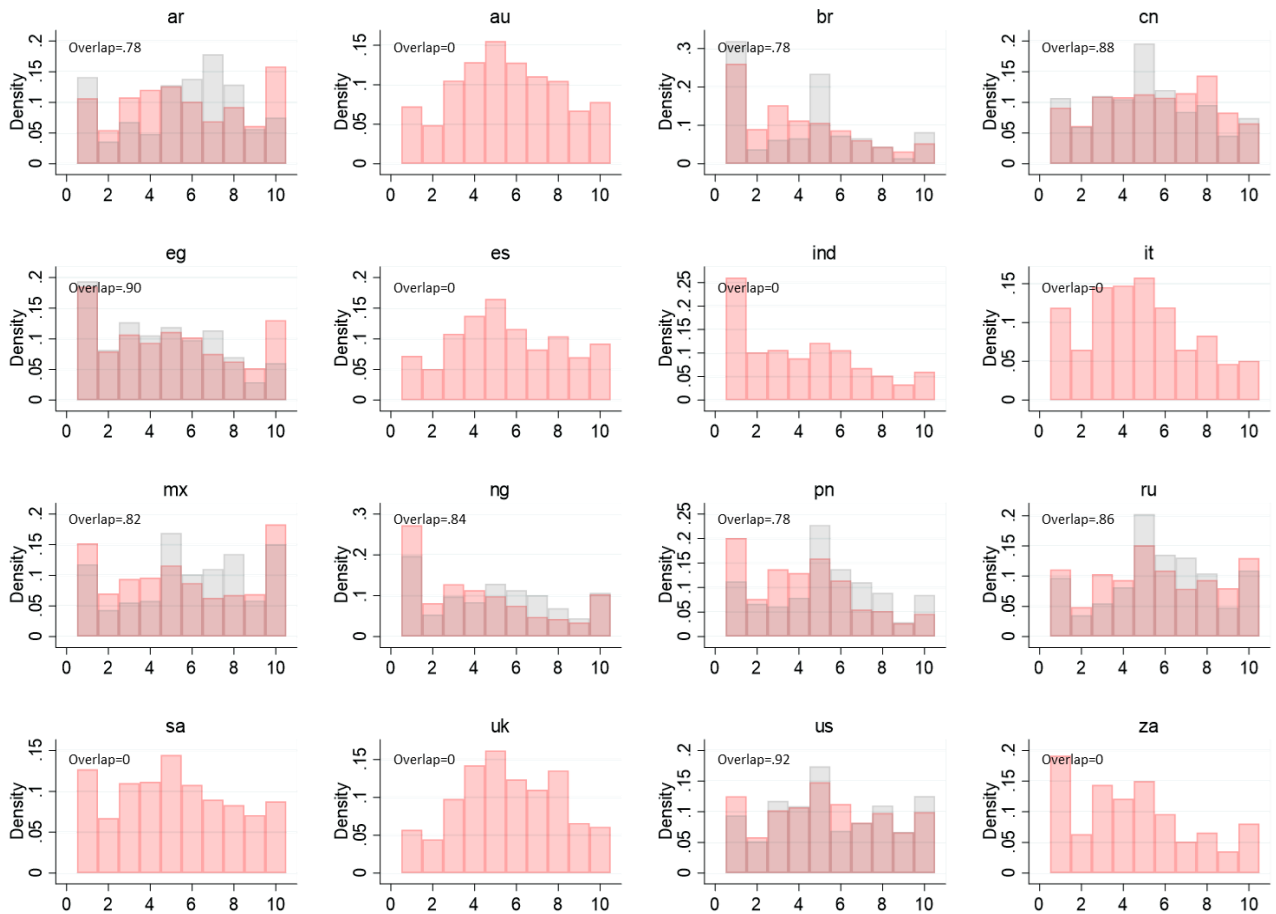


Fig. S4C. Distribution of responses by country to the World Values Survey (WVS) question 176 (How much do you agree or disagree with the statement that nowadays one often has trouble deciding which moral rules are the right ones to follow?). The bars in red depict our data, whereas the gray ones depict the most recent country values reported by the WVS. Each comparison also includes the percentage of overlap between the two samples. Certain WVS survey questions were not collected in certain countries by the WVS; for these countries, the WVS survey comparison is left blank.

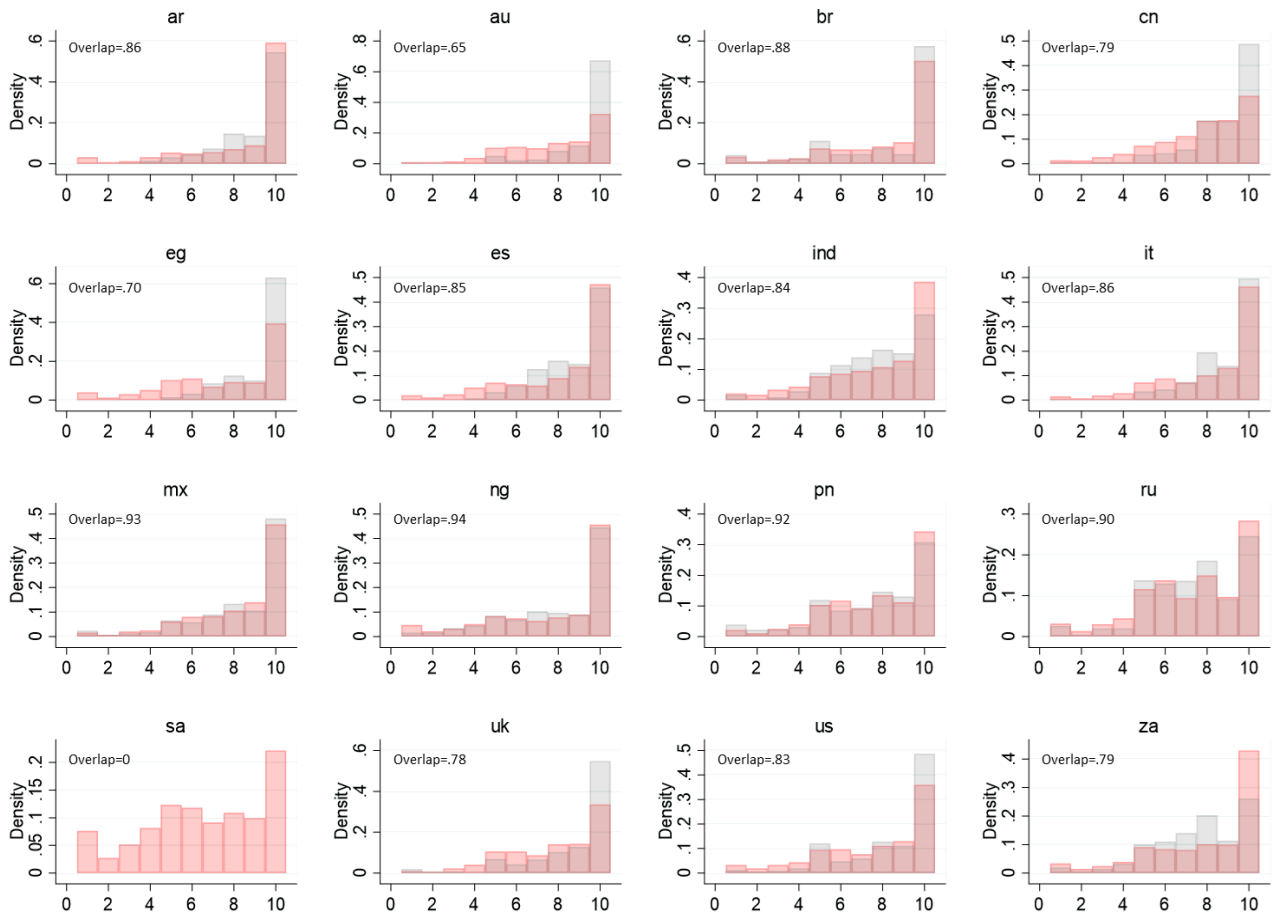


Fig. S4D. Distribution of responses by country to the World Values Survey (WVS) question 250 (How important is it for you to live in a country that is governed democratically?). The bars in red depict our data, whereas the gray ones depict the most recent country values reported by the WVS. Each comparison also includes the percentage of overlap between the two samples. Certain WVS survey questions were not collected in certain countries by the WVS; for these countries, the WVS survey comparison is left blank.

3. Supporting Analyses

3.1 Country-level predictors

Here we explore country-level differences using two of Hofstede's cultural dimensions[63]: power distance (how much a culture expects and accepts unequally distributed power) and individualism (how much a culture prioritizes personal interests over collective interests), as well as the most recent data available for the Corruption Perceptions Index[64] (Transparency International), Human Development Index[65] (United Nations), Global Freedom scores[66] (Freedom House), GDP per capita[67] (PPC 2019; The World Bank), and economic inequality[68] (Gini coefficient; the World Bank). In Table S6a, we show the results of calculating average discernment at the country-level and correlating with the various country-level variables (16 observations total). In Table 6b, we show the results of multi-level models that pool data from all countries and, for each country-level variable, we interact a veracity dummy, dummies for each treatment, and the country-level variable. We conduct one model for each country-level variable (using data from all conditions), and examine the three-way interactions between headline veracity, experimental condition, and country-level variable, which captures the extent to which the country-level variable explains moderates the treatment effect across countries. In Table 6c, we take a similar approach for individual differences in accuracy discernment (main text Figure 3). We run a multi-level model for each pair of individual-level variable and country-level variable (restricting to the Accuracy condition) and examine the 3-way interaction between headline veracity, the individual-level variable, and the country-level variable. This captures the extent to which the country-level variable moderates the relationship between the individual-level variable and accuracy discernment. These results are purely exploratory. Future work should investigate these relationships in greater detail, using a larger number of countries.

	Discernment in Accuracy condition	Discernment in Sharing condition	Difference in discernment
1) Gini coefficient	.248	.133	.106
	.3726	.6378	.7075
2) GDP per capita	.396	-.116	.441
	.1286	.6683	.0876
3) Power distance (Hofstede)	-.682	.064	-.652
	.0036	.8146	.0062
4) Individualism (Hofstede)	.645	-.309	.811
	.0070	.2438	.0001
5) Corruption Index	.501	-.221	.615
	.0479	.4105	.0112
6) Human Development Index (HDI)	.519	.109	.373
	.0395	.6882	.1546
7) Global Freedom score	.687	.054	.565
	.0033	.8436	.0225

Table S6A. Pearson's Correlations between country-level moderators and truth discernment. The first element of each row depicts the correlation, the second depicts the p-value. Numbers in black depict values where the corresponding p-value<0.05.

	2-way Inter.	Gini index	GDP pc PPP19	Power Dist.	Corrup. Index	HDI	Freedom Index	Individ. Index
Accuracy	**.118**	.007	**.029**	**-.041**	**.039**	**.024**	**.036**	**.052**
	(.011)	(.004)	**(.005)**	**(.005)**	**(.005)**	**(.005)**	**(.005)**	**(.006)**
	<.0001	.0990	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**
Prompt	**.023**	.006	.007	-.008	.010	.005	.009	.012
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
	<.0001	.0731	.0146	.0101	.0009	.0853	.0046	.0003
Tips	.013	.002	.001	-.003	.003	.001	.006	.004
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
	.0001	.6159	.7523	.3754	.4014	.7892	.0708	.2140

Table S6B. 3-way interactions between veracity, treatment condition, and country-specific variables (Share condition as holdout). First column shows 2-way interaction between veracity and treatment without any country-specific variables (to contextualize moderation effect sizes). Results are from multi-level models that pool data across all 16 countries (with random intercepts and slopes for both subject nested within country, and crossed with headline). The first element of each row depicts the coefficient, the second depicts the standard error, and the third the p-value (two-sided test). Numbers in black (bold) depict values where the corresponding p-value<0.05 (<.0001).

	2-way Inter.	Gini index	GDP pc PPP19	Power Dist.	Corrup. Index	HDI	Freedom Index	Individ. Index
2-way Inter.	-	.013	.023	-.039	.029	.030	.039	.037
		(.016)	(.016)	(.013)	(.015)	(.015)	(.013)	(.013)
		.4090	.1616	.0056	.0655	.0598	.0053	.0115
Pref. for Thinking	**.045**	-.005	.007	-.009	.009	.004	.004	.008
	(.004)	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)	(.003)
	<.0001	.0343	.0035	.0001	.0002	.1171	.1735	.0015
Thinking Perf.	**.039**	-.003	.004	-.003	.005	.000	.000	.005
	(.004)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
	<.0001	.2140	.0984	.1699	.0383	.9512	.9782	.0402
Attentiveness	**.046**	-.002	.003	-.006	.005	-.000	.007	.009
	(.003)	(.002)	(.002)	(.003)	(.002)	(.002)	(.002)	(.003)
	<.0001	.4959	.1694	.0177	.0452	.8717	.0030	.0007
Imp. of Accuracy	**.056**	-.000	**.012**	**-.012**	**.013**	.010	.009	**.013**
	(.004)	(.002)	**(.002)**	**(.003)**	**(.003)**	(.003)	(.002)	**(.003)**
	<.0001	.8850	**<.0001**	**<.0001**	**<.0001**	.0002	.0002	**<.0001**
Evidence / Cues	**.039**	-.004	**.011**	-.007	**.011**	.010	.008	.010
	(.004)	(.003)	**(.003)**	(.003)	**(.003)**	(.003)	(.003)	(.003)
	<.0001	.0788	**<.0001**	.0052	**<.0001**	.0001	.0036	.0005
Trust	.003	.003	-.004	.003	-.004	-.001	-.001	-.006
	(.003)	(.003)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
	.2920	.3026	.0732	.1894	.0717	.6987	.6942	.0210
Imp. of Democracy	**.054**	.001	**.013**	**-.016**	**.014**	.011	**.015**	**.018**
	(.004)	(.002)	**(.003)**	**(.003)**	**(.003)**	(.003)	**(.003)**	**(.003)**
	<.0001	.6550	**<.0001**	**<.0001**	**<.0001**	.0002	**<.0001**	**<.0001**
Ind. Responsibility	**-.015**	.006	-.003	-.000	-.001	.000	.004	.000
	(.003)	(.003)	(.002)	(.002)	(.002)	(.003)	(.002)	(.002)
	<.0001	.0174	.2773	.9107	.7103	.9347	.0749	.9730
Belief in God	-.013	.001	-.008	**.013**	**-.012**	-.007	**-.015**	**-.017**
	(.003)	(.002)	(.002)	**(.002)**	**(.002)**	(.002)	**(.002)**	**(.002)**
	.0002	.5910	.0002	**<.0001**	**<.0001**	.0011	**<.0001**	**<.0001**
Incentives / Equity	.003	-.001	-.007	.007	-.006	-.009	-.003	-.006
	(.002)	(.003)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
	.2823	.7595	.0055	.0022	.0088	.0007	.1695	.0106
Moral Relativism	.009	-.002	**.012**	**-.014**	**.011**	.012	**.010**	**.013**
	(.003)	(.003)	**(.003)**	**(.003)**	**(.003)**	(.003)	**(.002)**	**(.003)**
	.0028	.3442	**<.0001**	**<.0001**	**<.0001**	.0001	**<.0001**	**<.0001**
Urban	**.022**	.003	-.002	.003	-.005	-.002	.000	-.004
	(.003)	(.003)	(.002)	(.002)	(.002)	(.003)	(.002)	(.002)
	<.0001	.2430	.3310	.2357	.0241	.4982	.9814	.0767
Risk Attitude	**-.013**	.006	**-.015**	**.016**	**.017**	**-.012**	-.011	**-.019**
	(.003)	(.003)	**(.003)**	**(.003)**	**(.003)**	**(.003)**	(.003)	**(.003)**
	<.0001	.0318	**<.0001**	**<.0001**	**<.0001**	**<.0001**	.0002	**<.0001**
Minority	**-.020**	.010	-.009	.004	-.006	-.010	-.005	-.006
	(.003)	(.003)	(.003)	(.002)	(.002)	(.003)	(.002)	(.002)
	<.0001	.0003	.0003	.1118	.0219	.0005	.0381	.0064
Vaccine Likelihood	**.049**	-.001	**.014**	**-.015**	**.014**	**.016**	**.016**	**.017**
	(.006)	(.003)	**(.003)**	**(.003)**	**(.003)**	**(.003)**	**(.003)**	**(.003)**
	<.0001	.7909	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**
Vaccination Norm	**.032**	.000	**.014**	**-.017**	**.016**	**.015**	**.018**	**.017**
	(.004)	(.002)	**(.003)**	**(.003)**	**(.003)**	**(.003)**	**(.003)**	**(.003)**
	<.0001	.8470	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**

Table S6C. 3-way interactions between veracity, individual differences, and country-specific variables (Accuracy condition only). First column shows 2-way interaction between veracity and individual difference without any country-specific variable (to contextualize moderation effect sizes); first row shows 2-way interaction between veracity and country-level variables without any individual differences (as a complement to the correlations in Table S6a). Results are from multi-level models pool data across all 16 countries (with random intercepts and slopes for both subject nested within country, and crossed with headline). The first element of each row depicts the coefficient, the second depicts the standard error, and the third the p-value (two-sided test). Numbers in black (bold) depict values where the corresponding p-value<0.05 (<.0001).

3.2 Individual-level moderators and truth discernment

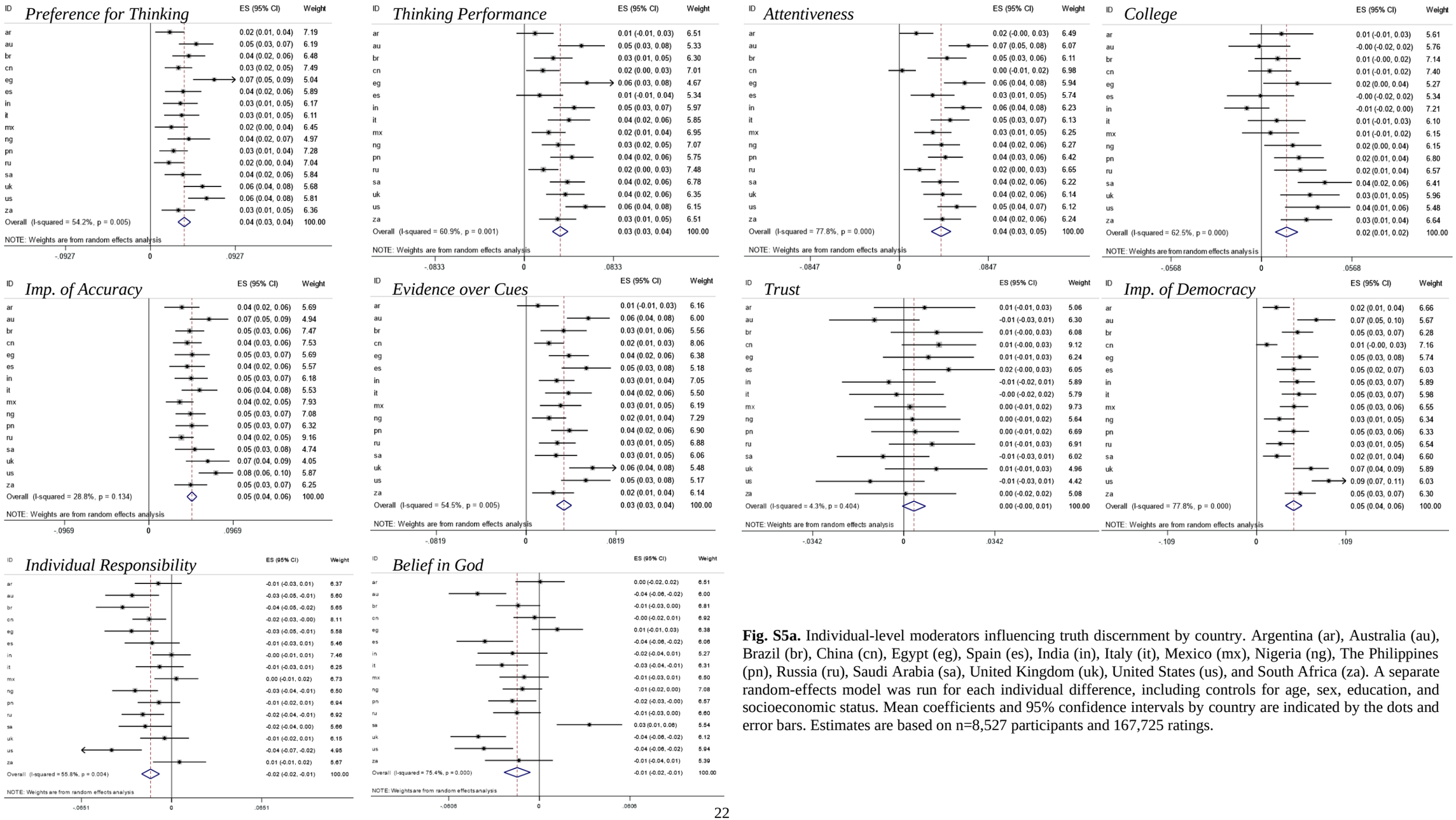


Fig. S5a. Individual-level moderators influencing truth discernment by country. Argentina (ar), Australia (au), Brazil (br), China (cn), Egypt (eg), Spain (es), India (in), Italy (it), Mexico (mx), Nigeria (ng), The Philippines (pn), Russia (ru), Saudi Arabia (sa), United Kingdom (uk), United States (us), and South Africa (za). A separate random-effects model was run for each individual difference, including controls for age, sex, education, and socioeconomic status. Mean coefficients and 95% confidence intervals by country are indicated by the dots and error bars. Estimates are based on n=8,527 participants and 167,725 ratings.

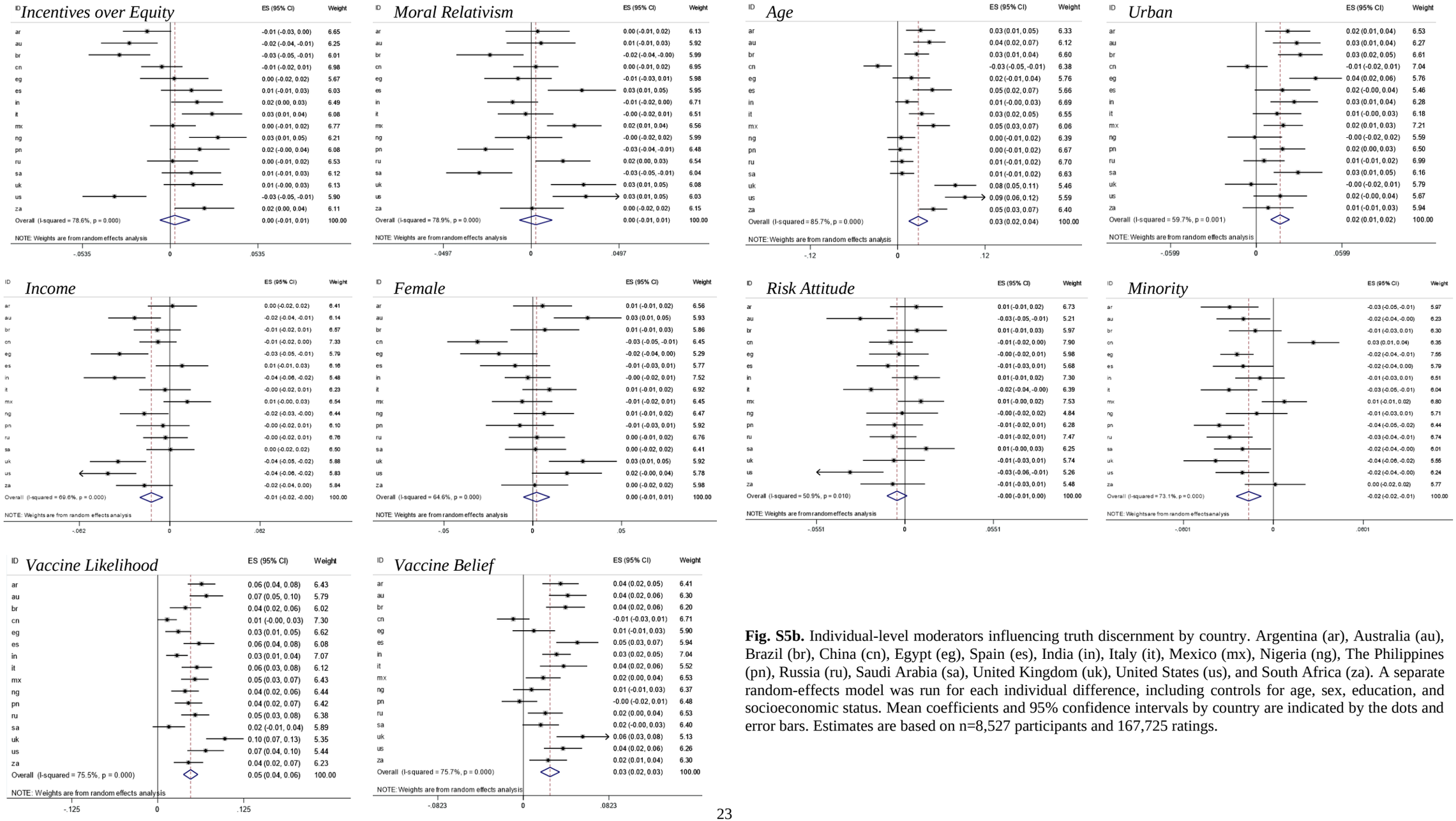


Fig. S5b. Individual-level moderators influencing truth discernment by country. Argentina (ar), Australia (au), Brazil (br), China (cn), Egypt (eg), Spain (es), India (in), Italy (it), Mexico (mx), Nigeria (ng), The Philippines (pn), Russia (ru), Saudi Arabia (sa), United Kingdom (uk), United States (us), and South Africa (za). A separate random-effects model was run for each individual difference, including controls for age, sex, education, and socioeconomic status. Mean coefficients and 95% confidence intervals by country are indicated by the dots and error bars. Estimates are based on n=8,527 participants and 167,725 ratings.

3.3 Moderators of truth discernment including controls and/or non-linear relationships

Table S7 column 1 demonstrates that the moderation relationships shown in the main text Figure 3 (including demographic controls and shown in Table S7 column 3) are similar when not including controls; shown is the coefficient on the interaction between a headline veracity dummy and the individual difference in question, with separate models run for each individual difference. Table S7 columns 2 and 4 investigate the effect of including quadratic terms for each individual difference (as well as the interaction between the quadratic term of the headline veracity dummy, which is our quantity of interest).

	Without controls			*With controls*		
Variable	*(1) Linear*	*(2) Quadratic*		*(3) Linear*	*(4) Quadratic*	
	Discernment	Linear Discernment	Quadratic Discernment	Discernment	Linear Discernment	Quadratic Discernment
Preference for Thinking	.044 [.000]	.044 [.000]	.007 [.016]	.037 [.000]	.037 [.000]	.010 [.000]
Thinking Performance	.038 [.000]	.037 [.000]	-.001 [.531]	.034 [.000]	.030 [.000]	.000 [.805]
Attentiveness	.046 [.000]	.041 [.000]	-.006 [.126]	.040 [.000]	.036 [.000]	-.005 [.206]
College	.016 [.000]	-	-	.016 [.000]	-	-
Imp. of Accuracy	.054 [.000]	.068 [.000]	.011 [.000]	.049 [.000]	.063 [.000]	.011 [.000]
Evidence over Cues	.037 [.000]	.052 [.000]	.012 [.000]	.034 [.000]	.053 [.000]	.017 [.000]
Imp. of Democracy	.053 [.000]	.076 [.000]	.023 [.000]	.045 [.000]	.070 [.000]	.024 [.000]
Individual Responsibility	-.014 [.000]	-.014 [.000]	.006 [.086]	-.015 [.000]	-.016 [.000]	.011 [.001]
Belief in God	-.012 [.015]	-.003 [.675]	.002 [.208]	-.015 [.001]	-.009 [.195]	.001 [.448]
Incentives over Equity	.004 [.452]	.008 [.150]	.010 [.001]	.003 [.544]	.009 [.085]	.014 [.000]
Moral Relativism	.009 [.146]	.007 [.222]	.015 [.000]	.003 [.583]	.001 [.914]	.016 [.000]
Age	.032 [.000]	.038 [.000]	-.012 [.000]	.028 [.000]	.033 [.000]	-.010 [.000]
Urban	.022 [.000]	.025 [.000]	.007 [.109]	.017 [.000]	.020 [.000]	.008 [.064]
SES	-.009 [.065]	-.012 [.014]	-.020 [.000]	-.012 [.002]	-.016 [.000]	-.018 [.000]
Female	.003 [.514]	-	-	.002 [.572]	-	-
Risk Attitude	-.012 [.036]	-.018 [.002]	-.007 [.005]	-.005 [.120]	-.007 [.041]	-.001 [.703]
Trust	.003 [.303]	.002 (.509]	-.004 [.175]	.004 [.089]	.005 [.038]	.002 [.222]
Minority	-.020 [.000]	-	-	-.016 [.000]	-	-
Vaccine Likelihood	.049 [.000]	.057 [.000]	.006 [.096]	.048 [.000]	.055 [.000]	.005 [.075]
Vaccine Norm	.032 [.000]	.022 [.000]	-.012 [.001]	.026 [.000]	.018 [.000]	-.010 [.002]

Table S7. The role of the focal individual differences on discernment. The first and the third models are linear and identify the role of the interaction of the headline being true and the value of the relevant variable (i.e., true x variable); the second and fourth models add a quadratic term and interaction to the model (i.e. variable^2 and true x variable^2) and identify the role of the linear and quadratic interactions (i.e. true x variable and true x variable^2). The first two models do not include any type of socio-demographic controls. The last two models include the following socio-demographic controls: age, sex, education, and socioeconomic status. P-values in brackets (two-sided test). The numbers in gray depict values where the p-value is greater than .05. For dichotomous variables, the associated values for the quadratic models are left blank.

Although many of the interactions between the quadratic term and headline veracity are significant, in most cases the magnitudes are small in relation to the interaction between the linear term and headline veracity. As a result, most individual differences do not show a substantial change in concavity with the relevant range of values. Figure S6 visualizes the overall interaction between the individual difference and headline veracity for the individual differences with significant quadratic interactions (using the models with controls). Exceptions include valuing incentives over equity, and moral relativism, which were not significant in the linear model but do show some evidence of a U-shaped relationship when including the quadratic term; and SES, for which the significant negative relationship in the linear model is found to be driven almost entirely by very high SES individuals.

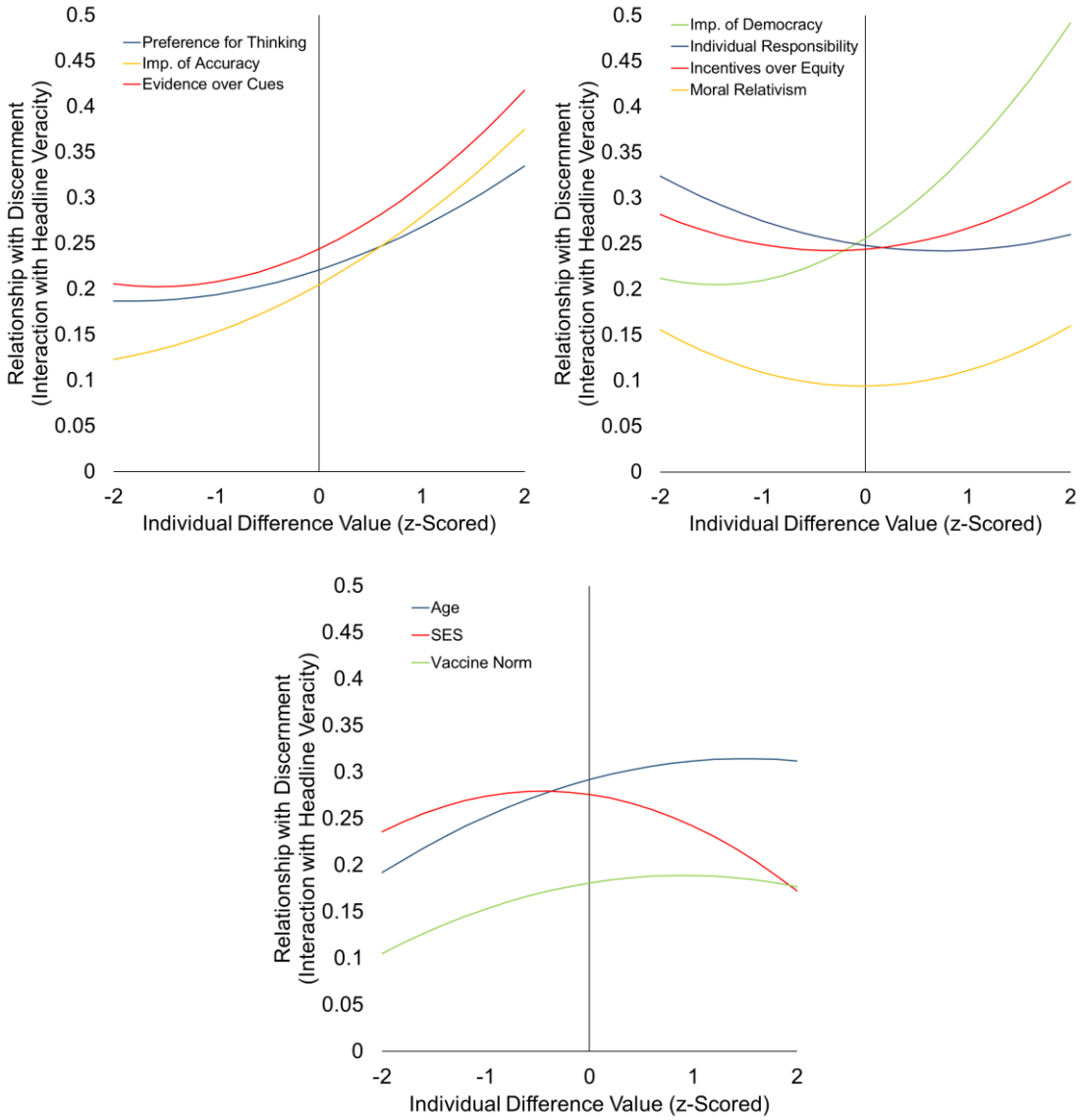


Fig. S6. Net coefficient on the interaction between individual difference and headline veracity (meta-analytic coefficient on true X difference + meta-analytic coefficient on true X difference^2) from the models in Table S6 column 3. Individual differences with no significant meta-analytic interaction between headline veracity and quadratic term are not included.

Finally, we shed further light on the non-linear effects by interaction headline veracity not only with the individual difference but also with the absolute value of the difference between the individual difference and the scale midpoint (capturing the *extremity* of the individual difference). Fig. S7 plots the coefficients for the interaction between headline

veracity and individual difference extremity, and shows that across nearly all Likert scale measures, more extreme responses (i.e. responses that are further from the scale midpoint) are associated with better discernment. (We focus on the Likert scales because extremity of response is not clearly defined for the non-Likert measures.) This observation accords with prior findings whereby, for example, Americans with more extreme ideological views are more reflective (i.e. score higher on the Cognitive Reflection Test) [69].

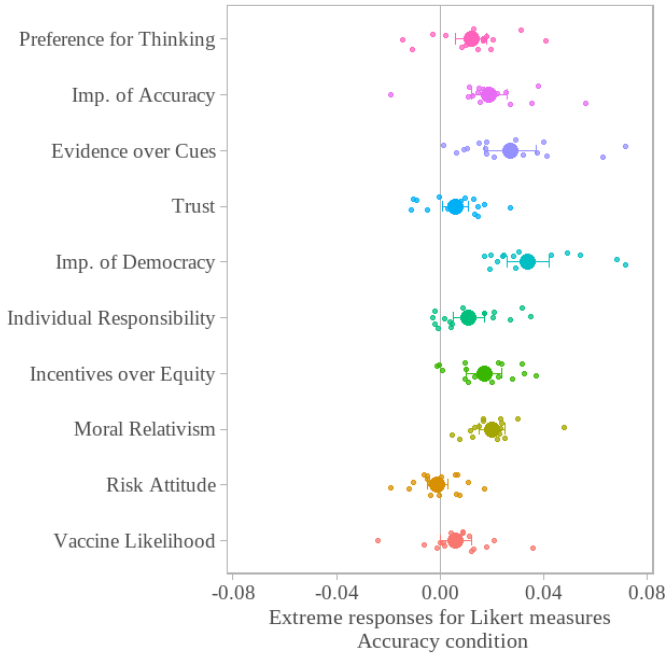
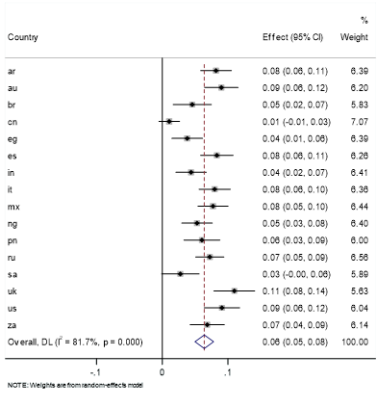


Fig. S7. Extreme responses for Likert measures in the accuracy condition. For each individual difference measure, shown is the coefficient of the interaction between headline veracity and the z-scored value of the absolute value of individual difference minus its midpoint in the corresponding Likert scale, when predicting perceived accuracy. Thus, the x-axis indicates the percentage point increase in accuracy discernment associated with a one standard deviation increase in the individual difference measure. The meta-analytic estimate and 95% confidence interval are indicated by the large dot and error bars; the smaller dots increase the estimate for each country. Estimates are based on n=8,527 participants and 167,725 ratings.

A)



B)

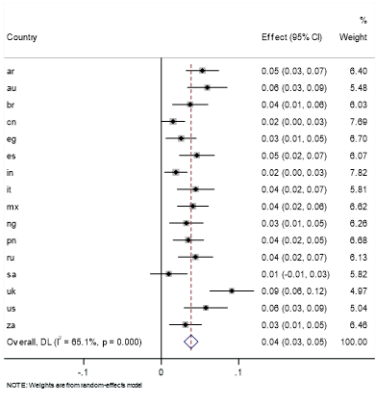


Fig. S8. Meta-analysis across countries of the correlations between truth discernment and COVID-19 vaccination intentions using A) vaccine-related false headlines and B) non-vaccine-related false headlines. Horizontal lines show meta-analytic mean estimates from random-effects models by country with 95% confidence intervals. Estimates are based on n=8,527 participants and 167,725 ratings.

3.4 Sharing intentions for true and false headlines in Sharing condition, by country

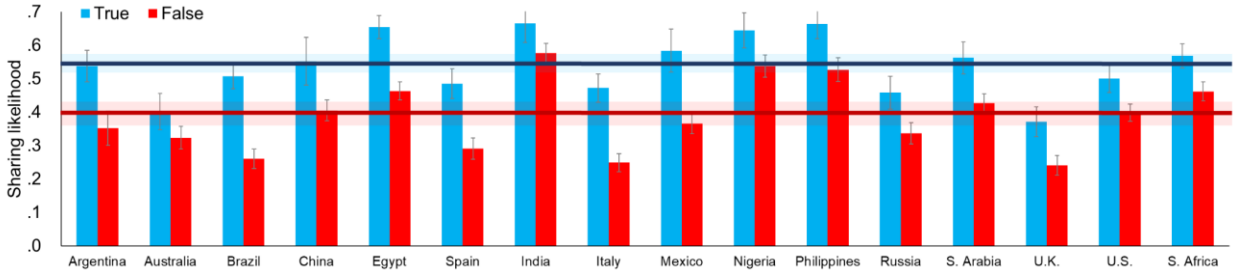


Fig. S9 Average likelihood of sharing true (blue) and false (red) headlines by country (error bars indicate 95% confidence intervals); sorted by average sharing discernment. Horizontal lines show meta-analytic mean estimates with 95% confidence intervals. $n_{Participants}$=8,631; $n_{Ratings}$=170,511.

3.5 Individual difference predictors of baseline sharing discernment in Sharing condition

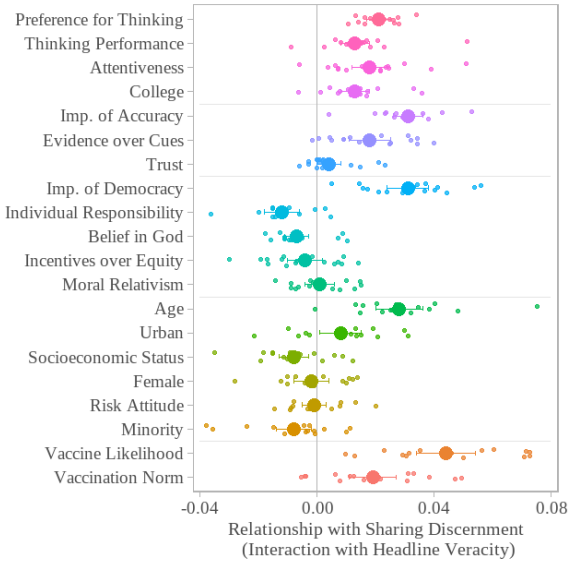


Fig. S10. Individual-level moderators of sharing discernment (sharing condition). $n_{Participants}$=8,631; $n_{Ratings}$=170,511.

3.6 Effect of the Prompt and Tips conditions on sharing of true and false headlines

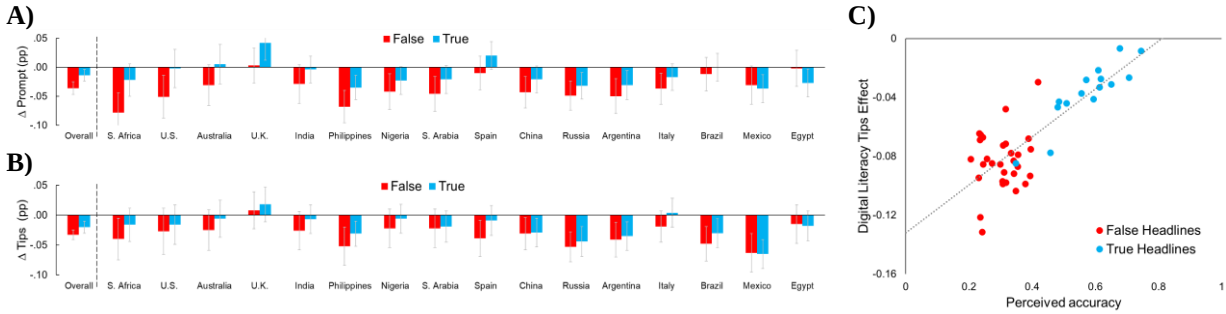


Fig. S11. Mean change in sharing intentions of false and true headlines in A) Prompt and B) Tips conditions relative to baseline Sharing condition. Error bars show 95% confidence intervals; C) Variation in effect of Tips condition across items. $n_{Participants}$=34,286; $n_{Ratings}$=676,605.

3.7 Moderators of Prompt and Tips effects on sharing discernment

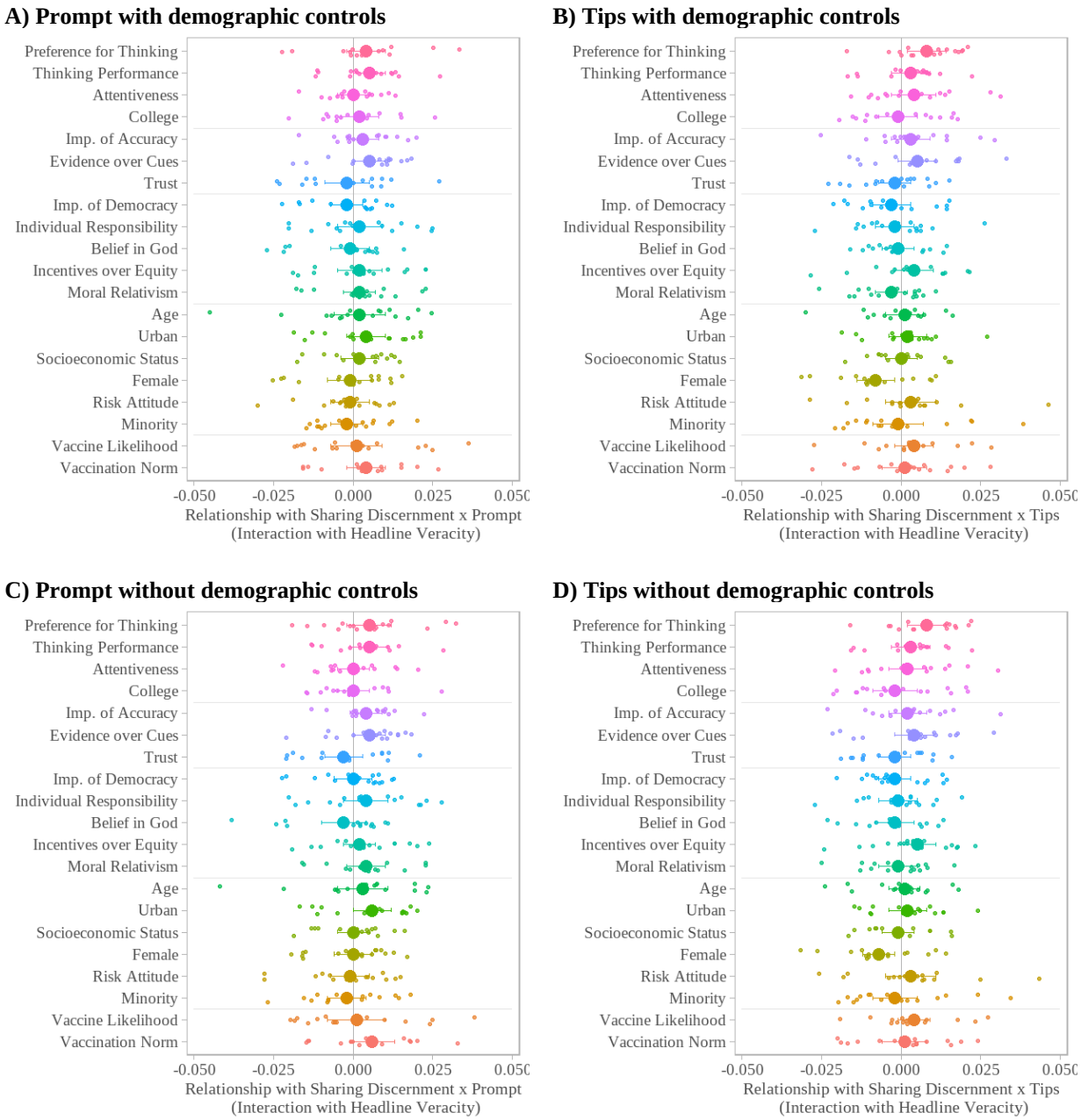
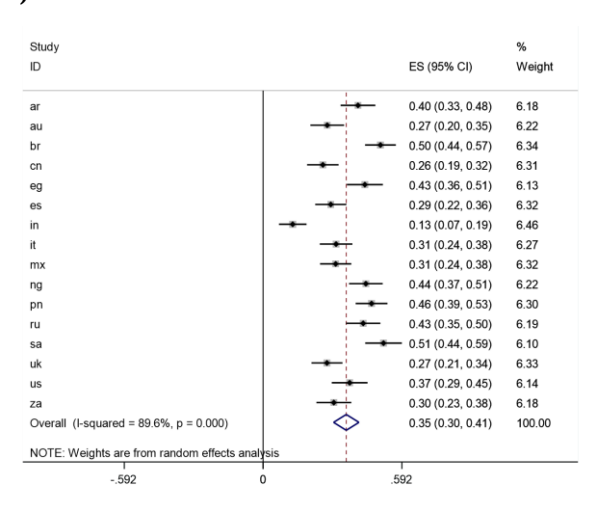


Fig. S12. Three-way interactions between individual differences, veracity and prompt (left; A and C) and tips (right; B and D). Baseline sharing as holdout. Top panel (A & B) includes demographic controls (along with their interactions with veracity) for age, sex, education (college degree), and income; bottom panel does not (C & D). The meta-analytics estimate and 95% confidence interval are indicated by the large dot and error bars; the smaller dots show the estimate for each country. $n_{Participants}$=34,286; $n_{Ratings}$=676,605.

3.8 Helpfulness and likeability of the Prompt and Tips conditions

A)



B)

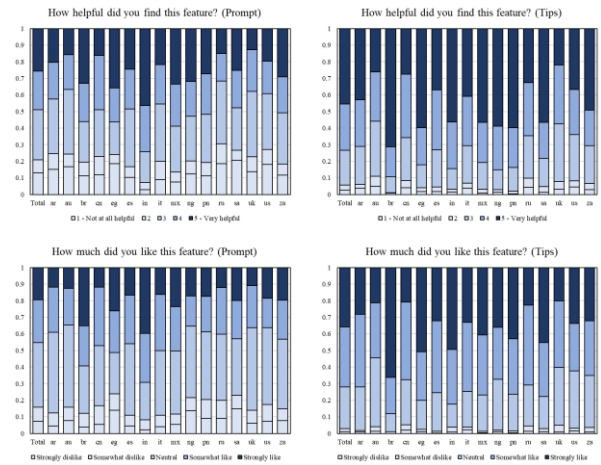
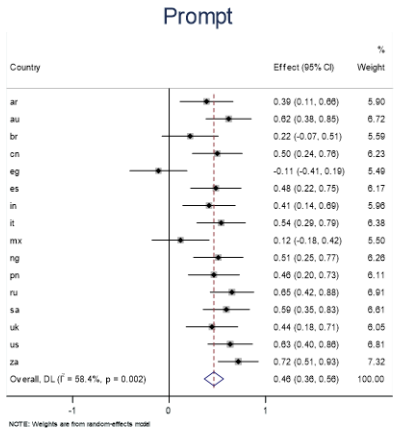


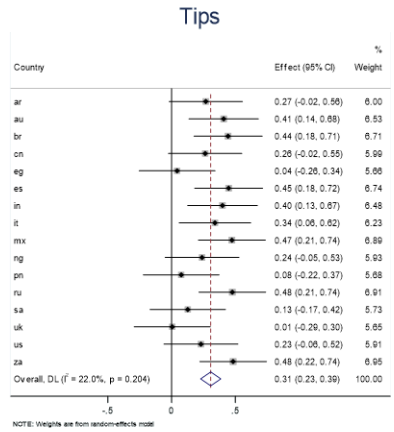
Fig. S13. A) Random-effects meta-analysis of the mean difference in helpfulness rating of the Tips interaction relative to the Prompt intervention (positive values imply higher perceived helpfulness of Tips). Mean coefficients and 95% confidence intervals by country are indicated by the dots and error bars. $n_{Participants}$=17,128; $n_{Ratings}$=338,369. B) Full distributions of helpful and likeability ratings for each intervention and country. N=15,864.

3.9 Item analysis: Perceived accuracy as a predictor of treatment effects

A)



B)



C)

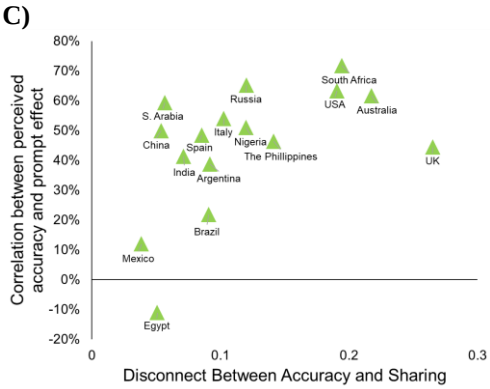


Fig. S14. Random-effects meta-analysis across countries of the correlations between perceived accuracy and A) Prompt effect and B) Tips effect; mean coefficients and 95% confidence intervals by country are indicated by the dots and error bars. C) correlation between perceived accuracy and Prompt effect as a function of the disconnect between accuracy and sharing (difference in discernment in Accuracy relative to Sharing conditions). For A-C, $n_{Participants}$=34,286; $n_{Ratings}$=676,605.

As pre-registered, we conduct item analyses using subjective accuracy (shown in aggregate in main text Figure 5c) separately for each country. First, for each headline, we calculated the average perceived accuracy in the *Accuracy* condition. Then, for each headline, we calculated the treatment effect (i.e., difference in sharing between the control and treatment, divided by the control) for the *Prompt* condition and the *Tips* condition. Finally, we calculated the correlation between perceived accuracy and each of the two (country-specific and z-scored) treatment effects. We then meta-analyzed these two correlation coefficients across countries to test whether there are significant positive correlations. We find a positive effect for *Prompt* (meta-analytic estimate, r=0.464, z=9.16, p<0.001; Figure S14a), and for *Tips* (meta-analytic estimate, r=0.307, z=7.57, p<0.001; Fig S14b).

We find significant heterogeneity across countries in the magnitude of correlation between perceived accuracy and the effect of *Prompt* (χ^2=36.02, p=0.002), but not in *Tips* (χ^2=19.23, p=0.204). Thus, we also examine how the correlations between perceived accuracy and prompt effect vary based on the disconnect between sharing and accuracy discernment (Fig S14c; subjective accuracy analog to Fig 5b in the main text). As with the relationship between the prompt effect and objective accuracy, we see that the extent to which the prompt reduces sharing of content that is perceived as inaccurate is greater in countries where there is more of a baseline disconnect between perceived accuracy and sharing intentions.

3.10Ratings from small groups of laypeople with and without a bachelor's degree

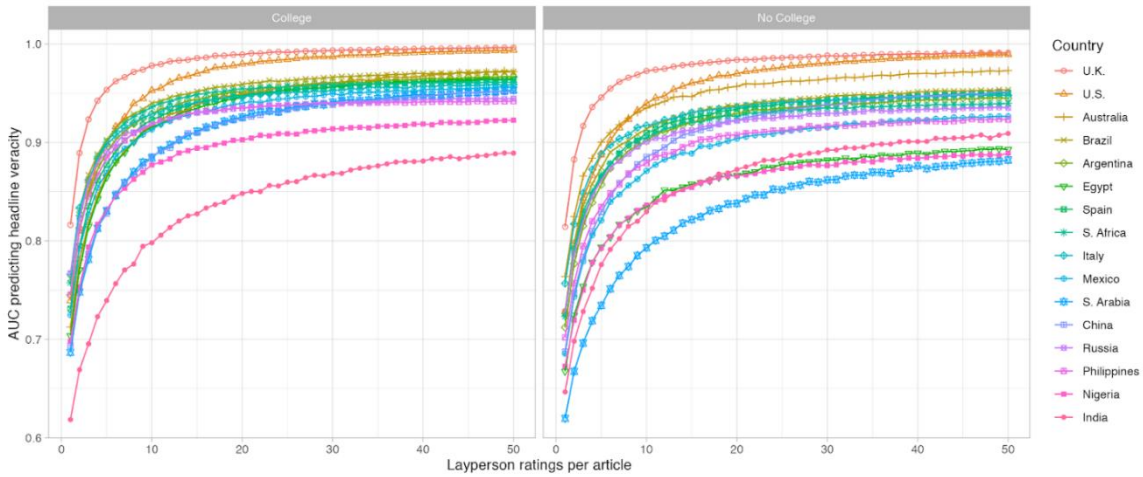


Fig. S15. Ratings from small groups of laypeople with and without a bachelor's degree can reliably identify misinformation. Area under the curve (AUC) when predicting headline veracity using the average rating of a crowd of *k* layperson respondents, for each country, by education attainment (college degree or more versus less than a college degree).

4. Supplemental References

63. Hofstede, G. & Bond, M. H. Hofstede's Culture Dimensions: An Independent Validation Using Rokeach's Value Survey. *J. Cross. Cult. Psychol.* **15**, 417–433 (2016).
64. Transparency International. Corruption Perceptions Index. (2021). Available at: https://www.transparency.org/en/cpi/2021. (Accessed: 11/2/2022)
65. United Nations. Human Development Index. *Human Development Reports* (2022).
66. Freedom House. Global Freedom Scores. *Countries and Territories* (2022). Available at: https://freedomhouse.org/countries/freedom-world/scores. (Accessed: 11/2/2022)
67. The World Bank. GDP per capita. (2022). Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?end=2019&start=1990. (Accessed: 11/2/2022)
68. The World Bank. Gini index. (2022). Available at: https://data.worldbank.org/indicator/si.pov.gini?most_recent_value_desc=false. (Accessed: 11/2/2022)
69. Pennycook, G. & Rand, D. G. Cognitive Reflection and the 2016 U.S. Presidential Election. *Personal. Soc. Psychol. Bull.* **45**, (2019).