Supplementary information for

Ability to detect fake news predicts geographical variation in COVID-19 vaccine uptake

Sahil Loomba, Rakoen Maertens, Jon Roozenbeek, Friedrich M. Götz, Sander van der Linden, Alexandre de Figueiredo

Correspondence to: Sahil Loomba, s.loomba18@imperial.ac.uk

Supplementary results

Misinformation susceptibility and first dose uptake

In accordance with the preregistered analyses [\[S1\]](#page-88-1), a secondary set of vaccine uptake models was also inferred, with the outcome being a COVID-19 vaccine's *first* dose uptake rates in England and Scotland as of 1 October 2021; see Fig. [S19.](#page-42-0) While the posterior estimates remain similar to those in the primary vaccine uptake models (Fig. 1), the 95% HPDIs for the coefficient of fake news detection ability score M_f^{μ} f_f^{μ} overlap with the ROPE: β_M^V $\frac{N_{\mu}}{M_{f}^{\mu}}=0.25\, [0.02, 0.47]$ and $\beta_{\scriptscriptstyle M}^V$ $M_f^V = 0.22$ [0.01, 0.43], both before and after controlling for spatial effects, respectively. That is, the practical significance of the effect of M_f^{μ} f ^{μ} on first dose uptake rates as of October 2021 cannot be determined. However, the region of overlap reduced considerably upon considering first dose uptake rates as of 1 July 2021, i.e. three months prior (consistent with the recommended gap between the first and second doses [\[S2\]](#page-88-2)), to β_{μ}^{V} $\frac{dN_f^{\mu}}{dt} = 0.24 \, [0.05, 0.44] \; {\rm and} \;$ $\beta^V_{\scriptscriptstyle{M}}$ $M_f^V = 0.23$ [0.04, 0.42], before and after controlling for spatial effects, respectively.

Interpreting MIST scores

One can interpret the region-level model estimates of MIST scores by considering reference scores under a null model of random guessing. If an individual is equally likely to respond to a news item with "real" or "fake", then it can be shown that the expectations of real (M_r^{μ}) and fake $(M_f^{\mu}$ f) news detection ability scores are given by 5, of veracity discernment ability score (M_v^{μ}) by 10, and of distrust (M_d^{μ}) d_d^{μ}) and naivety (M_n^{μ}) bias scores by ≈ 0.88 ; see *Methods: Poststratifying MIST scores*. Whether the 95% HPDI lies below, includes, or lies above the reference values, determines if a region performs worse than, close to, or better than random guessing, respectively. Only a single region is observed to perform worse than random guessing, and across just one score of real news detection ability: M_r^{μ} [UKG38 (Walsall)] = 4.39 [4.07, 4.77]. Of the 149 regions across the UK, 34 perform close to random guessing for the real news detection ability score M_r^{μ} ; see Table [S11.](#page-66-0) Only one region performs close to random guessing for both bias scores M_d^{μ} $\frac{\mu}{d}$ and M_n^{μ} : M_d^{μ} $\frac{\mu}{d}$ [UKN16 (Fermanagh and Omagh)] $= 0.84$ [0.47, 1.21] and M_n^{μ} [UKN16 (Fermanagh and Omagh)] = 0.86 [0.49, 1.27]. Consequently, all regions perform better than random guessing at detecting fake news and veracity discernment.

Determinants of abilities to detect real and fake news

From the inferred social-IRT models, Fig. 3 shows the posterior mean of the contribution β_{Cov} [C] (or $\underline{\beta}_{\text{Cov}}$ [C]) of belonging to a group C along a socio-demographic covariate COV, relative to its reference group, to an individual's log-odds of correctly detecting a real (or fake) news headline with unit discrimination. All else held constant, $\beta_{\text{Cov}}[C] > 0$ ($\beta_{\text{Cov}}[C] < 0$) implies that group C has a higher (lower) ability to correctly identify a real headline, relative to the reference group. Groups with a log-odds credibly different from 0 relative to the reference group, based on 95% HPDIs, are listed below.

Relative to 18–24 year olds (the baseline group for age), older age groups get progressively better at detecting both real— β_{AGE} [25–34] = 0.21 [0.06,0.36], β_{AGE} [35-44] =

 $0.46\,[0.23,0.7], \quad \quad \beta_{\rm AGE}\,[45\hbox{-}54] \;\;=\;\; 0.88\,[0.5,1.32], \quad \quad \beta_{\rm AGE}\,[55\hbox{-}64] \;\;=\;\; 0.97\,[0.55,1.45],$ β_{AGE} [65+] = 0.72 [0.38, 1.09]—and fake news— $\underline{\beta}_{\text{Age}}$ [25–34] = 0.37 [0.19, 0.57], $\frac{\beta_{\rm AGE}}{\beta_{\rm AGE}}[35\text{-}44] \;\; = \;\; 1.06 \,[0.6, 1.56], \quad \frac{\beta_{\rm AGE}}{\beta_{\rm AGE}}[45\text{-}54] \;\; = \;\; 1.78 \,[1.04, 2.63], \quad \frac{\beta_{\rm AGE}}{\beta_{\rm AGE}}[55\text{-}64] \;\; = \;\; 1.06 \,[0.6, 1.56],$ 2.49 [1.43, 3.65], β_{AGE} [65+] = 3.0 [1.74, 4.4]. Evidently, the relationship is much stronger for fake news detection ability. Relative to those without an education qualification, those with higher qualifications get progressively better at detecting both real— β_{EDU} [Level-1] = 0.51 [0.24, 0.78], β_{EDU} [Level-2] = 0.9 [0.47, 1.32], β_{EDU} [Level-3] = 1.2 [0.64, 1.75], β_{EDU} [Level-4] = 1.72 [0.93, 2.49]—and fake news— $\frac{\beta_{\rm EDU}[\textsf{Level-1}] = 0.55[0.28, 0.86], \;\;\underline{\beta_{\rm EDU}[\textsf{Level-2}]} = 0.94[0.53, 1.41], \;\;\underline{\beta_{\rm EDU}[\textsf{Level-3}]} = 0.095[0.53, 1.41],$ 1.62 [0.92, 2.39], $\underline{\beta}_{\text{EDU}}$ [Level-4] = 1.73 [1.0, 2.56]. The gains in ability with education level are more similar between real and fake news detection abilities. Relative to females, males are better at real news detection: β_{GEN} [Male] = 0.23 [0.12, 0.33]. Relative to being white, ethnic minority groups have a lower ability to discern real and fake news: β_{ETH} [Other] = -0.25 [-0.4, -0.1], $\underline{\beta}_{\text{ETH}}$ [Other] = -0.57 [-0.84, -0.31]. Relative to being Christian, being atheist implies enhanced abilities for both real and fake news detection: β_{REL} [Atheist] = 0.86 [0.5, 1.27], β_{REL} [Atheist] = 0.81 [0.46, 1.2] while other religious groups have a lower ability to discern both real and fake news— $\beta_{\rm REL}$ [Other] = -0.13 [-0.25, -0.03], $\beta_{\rm REL}$ [Other] = -0.14 [-0.25, -0.03]. Students and retired groups have a better ability to discern real and fake news than those who are employed— β_{EMP} [Student] = 0.59 [0.29, 0.9], β_{EMP} [Retired] = 0.36 [0.16, 0.55], $\underline{\beta}_{\text{EMP}}$ [Student] $= 0.37$ $[0.17, 0.59]$, $\underline{\beta}_{\text{EMP}}$ [Retired] $= 0.62$ $[0.31, 0.95]$.

The social-IRT models accounted for regional fixed effects; all regional covariates were standardized to zero mean and unit standard deviation, and those with a positively skewed distribution—population density, income per head, and proportion of higher degree holders were log-transformed. Fig. 3 shows the posterior mean of the effect β_{Cov} (or β_{Cov}) of increasing the regional covariate COV by unit standard deviation on the log-odds of an individual, who resides in that region, to correctly detect a real (or fake) news headline with unit discrimination. Those residing in densely-populated regions are worse at detecting fake news $(\underline{\beta}_{\text{Pop}} = -0.13[-0.26, -0.01])$, while those living in regions with more higher degree holders fare better at detecting real news ($\beta_{DEG} = 0.18[0.04, 0.33]$).

The social-IRT models additionally accounted for regional random effects, whose structured component encoded for the underlying social network structure. Fig. [S7](#page-30-0) maps the structured, unstructured, and total random effects for real and fake news detection abilities in the UK. The model parameters $\rho \in [0,1], \rho \in [0,1]$ measure the proportion of variance in the total effect explained by the structured component [\[S3\]](#page-88-3) for real and fake news detection abilities, respectively. Their posterior estimates are similar and credibly larger than $0.5-\rho = 0.94[0.79,1.0]$ and $\rho = 0.92[0.7, 1.0]$ —indicating that social connectivity explains a majority of residual variance and that information spillovers may play an important role in real and fake news detection abilities.

For England, Scotland, and Wales—comprising the island of Great Britain (GB) a secondary set of social-IRT models was inferred with an expanded covariate set consisting of an additional individual covariate of income, a finer partitioning of two individual covariates of ethnicity and religious affiliation, and two additional regional covariates; see Table [S2.](#page-48-0) Fig. [S8](#page-31-0) shows the posterior estimates of corresponding model parameters which tend to agree with estimates for the entire UK shown in Fig. 3. This secondary analysis reveals that relative to those earning less than *£*25,000 a year, individuals with higher incomes are incrementally (and comparably) better at detecting both real— β_{INC} [Level-1] = 0.2 [0.09, 0.32], β_{INC} [Level-2] = 0.45 [0.27, 0.66]—and fake news— $\underline{\beta}_{\text{INC}}$ [Level-1] = 0.27 [0.13, 0.43], $\underline{\beta}_{\text{INC}}$ [Level-2] = 0.43 [0.24, 0.66]. Amongst the ethnic minority groups, while each of them have a lower fake news detection ability relative to being white— $\underline{\beta}_{\text{ETH}}$ [Asian] = -0.4[-0.65, -0.18], $\underline{\beta}_{\text{ETH}}$ [Black] = -0.9[-1.32, -0.48], β_{ETH} [Mixed] = -0.84 [-1.28, -0.45], β_{ETH} [Other] = -0.32 [-0.69, -0.03]—those with an Asian or mixed ethnic background appear to also have lower real news detection abilities: β_{ETH} [Asian] = -0.56 [-0.83, -0.3], β_{ETH} [Mixed] = -0.25 [-0.51, -0.02]. Amongst the religious minorities, being Muslim seems to imply a lower ability to detect fake news relative to being Christian: $\underline{\beta}_{\text{REL}}$ [Muslim] = -0.57 [-0.88, -0.28]. Individuals from regions that voted more favorably to remain in the EU had higher real news detection abilities, at $\beta_{\text{VOT}} = 0.21$ [0.04, 0.4], which appears to substitute for the effect of being from a region with a larger proportion of higher degree holders: $\beta_{DEG} = 0.0[-0.21, 0.23]$. It is noteworthy that the structured component encoding social network structure remains strong and similar for both abilities when considering just GB: $\rho = 0.87 \left[0.56, 1.0 \right]$ and $\rho = 0.86 \left[0.52, 1.0 \right]$. The resulting posterior distributions of poststratified regional MIST scores for GB under the secondary social-IRT models were very similar to those from the primary social-IRT models; see Fig. [S11](#page-34-0) for more details.

Determinants of vaccine uptake

Fig. 1 shows the posterior estimates of the standardized regression coefficients β_{Cov}^V which measures the effect of increasing the regional covariate COV by unit standard deviation on increase in a COVID-19 vaccine's second dose uptake rates in England and Scotland as of 1 October 2021, in units of standard deviation of regional vaccine uptake rates. Regional determinants of vaccine uptake rates with coefficients credibly and consistently different from 0 based on 95% HPDIs are enlisted below. Population density $(\beta_{\text{Pop}}^V = -0.25[-0.47, -0.04])$ and unemployment rate ($\beta_{\text{UNE}}^{V} = -0.16[-0.27, -0.06]$) are negatively associated with second dose uptake, while the proportion of population aged 60 or older is positively associated with it $(\beta_{\text{AGE}}^V = 0.31 [0.07, 0.53])$, while controlling for spatial effects. See Table [S15](#page-78-0) for the full set of posterior estimates.

Fig. 2 shows the posterior mean of the contribution β_{Cov}^{ν} [C] of belonging to a group C along a socio-demographic covariate COV, relative to its reference group, to an individual's log-odds of being vaccinated. All else held constant, $\beta_{\text{Cov}}^{\nu}[\textbf{C}] > 0$ ($\beta_{\text{Cov}}^{\nu}[\textbf{C}] < 0$) implies that group C has a higher (lower) probability to be vaccinated, relative to the reference group. Groups with a log-odds credibly different from 0 relative to the reference group, based on 95% HPDIs, are listed below.

Relative to 18–24 year olds, older age groups are progressively more likely to get vaccinated— β_{AGE}^{ν} [45-54] = 0.37 [0.09, 0.64], β_{AGE}^{ν} [55-64] = 1.09 [0.78, 1.4], β_{AGE}^{ν} [65+] = 1.36[0.93,1.79]. Relative to being white, being black is associated with a lower probability to be vaccinated: β_{ETH}^{ν} [Black] = -0.74 [-1.11, -0.37]. Those who are unemployed or unable to work have a lower probability to be vaccinated than those who are employed— $\beta_{\rm Emp}^{\nu} \, [{\sf Unemployed}] \; = \; -0.34 \, [-0.63, -0.07], \;\;\; \beta_{\rm EMP}^{\nu} \, [{\sf Other}] \; = \; -0.3 \, [-0.5, -0.08] - {\rm while}$ those who are retired have a relatively higher probability: β_{EMP}^{ν} [Retired] = 0.52 [0.14, 0.89]. We also observe some interesting differences in determinants of vaccine uptake from the determinants of misinformation susceptibility; see *[Supplementary results:](#page-1-0) [Determinants of abilities](#page-1-3)* *[to detect real and fake news](#page-1-3)*. Notably, we do not observe a credible trend in vaccination status with education levels, unlike the trend for real and fake news detection abilities, and atheists are less likely to be vaccinated relative to Christians $(\beta_{\text{REL}}^{\nu}$ [Atheist] = $-0.35[-0.52, -0.18]$) whereas they are better at detecting both real and fake news.

Finally, unlike for misinformation susceptibility, we do not find evidence for social connectivity explaining the residual variance in vaccination status, after controlling for misinformation susceptibility, individual- and region-level covariates. That is, the model parameter $\rho^v \in [0,1]$, which measures the proportion of residual variance explained by the structured component [\[S3\]](#page-88-3) that encodes social connectivity volumes, has a posterior that is not credibly different from 0.5: $\rho^{\nu} = 0.48$ [0.0, 0.99]. Together with the results that the ability to detect fake news predicts vaccine uptake—see *Results: Ability to detect fake news predicts COVID-19 vaccine uptake*—and that social network structure explains residual variance in fake news detection abilities—see *[Supplementary results:](#page-1-0) [Determinants of abilities to detect real and fake news](#page-1-3)*—this finding suggests that if vaccination status co-varies on social networks, then this network effect may be mediated by misinformation susceptibility.

Supplementary methods

Individual social connectivity generates regional spatial structure

As previously motivated, we assume that any residual correlations in the ability to detect real or fake news arise due to information flows, mediated by social connections at the individual level. In this section, we show how assuming correlations on an *individual*-level social network asymptotically translates to *region*-level spatial structure, that can be readily supplied by publicly available data on social connection volumes between regions [\[S4\]](#page-88-4).

ICAR prior at the individual level Let N be the total population size of the country, and φ be a vector of size *N* encoding individual-level random effects. Let A be the individual-level symmetric adjacency matrix of size $N \times N$, i.e. $A_{ij} = 1$ if individual *i* is socially connected with individual *j* and vice-versa ($i \neq j$, and the network is undirected) and $A_{ij} = 0$ otherwise. Analogous to Eq. 5, define a conditional probability distribution for φ that penalizes the average sum of squared differences in random effects of an individual with their social connections:

$$
\mathbb{P}(\varphi|\mathbf{A}) \propto \exp\left(-\frac{s_1}{2N} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\varphi_i - \varphi_j\right)^2 \mathbf{A}_{ij}\right)
$$

= $s_2 \prod_{i=1}^{N} \prod_{j=i+1}^{N} \Phi_{ij}^{\mathbf{A}_{ij}},$ (S1)

where

$$
\Phi_{ij} \triangleq \exp\left(-\frac{s_1}{2N}\left(\varphi_i - \varphi_j\right)^2\right),\tag{S2}
$$

and $s_1 > 0$, $s_2 > 0$ are positive constants. This is equivalent to assuming an intrinsic conditional autoregressive (ICAR [\[S5\]](#page-88-5)) prior at the individual level. Since an individual's social connections are expected to be bounded and small compared to the large population size *N*, what we call the "sparsity" assumption, penalizing the *average* ensures that the density in Eq. S1 remains invariant to population scaling.

Statistical network model If the social network structure A is known, then the prior on φ is completely specified by Eqs. S1 and S2. However, we do not observe the true network structure, i.e. A is latent, and consider the distribution of individual random effects after marginalization under an assumed statistical network model [\[S6\]](#page-88-6). In particular, we assume that conditioned on some *D*-dimensional space encoded by a matrix Θ of size $N \times D$, social connections between individuals *i*, *j* form independently from connections between individuals \tilde{i} , \tilde{j} ($i \neq j$, $\tilde{i} \neq \tilde{j}$, and $j \neq j$:

$$
\mathbf{A}_{ij} \perp \mathbf{A}_{\tilde{i}\tilde{j}} | \Theta_i, \Theta_j, \Theta_{\tilde{i}}, \Theta_{\tilde{j}},
$$
 (S3)

as commonly done in statistical modeling of social networks [\[S7\]](#page-88-7): see Fig. [S3.](#page-26-0) Since φ is defined to capture network structural effects, it implies that any influence of Θ on φ is completely mediated by the adjacency structure:

$$
\Theta \perp \varphi \mid A. \tag{S4}
$$

Network sparsity was encoded by assuming that the probability of a connection between any two individuals is small, scaling as:

$$
\mathbb{P}\left(\mathbf{A}_{ij}=1\,\big|\,\mathbf{\Theta}_i,\mathbf{\Theta}_j\right)=\mathscr{O}\left(N^{-1}\right),\tag{S5}
$$

that implies asymptotically bounded number of social connections per individual for large *N* [\[S8\]](#page-88-8), as assumed above. Undirectedness also implies symmetry of probabilities of connection:

$$
\mathbb{P}\left(\mathbf{A}_{ij}=1\,|\,\Theta_i,\Theta_j\right)=\mathbb{P}\left(\mathbf{A}_{ji}=1\,|\,\Theta_j,\Theta_i\right).
$$
\n(S6)

Let $\mathscr A$ be the space of all adjacency matrices. The distribution of individual effects φ conditioned on Θ can be written by marginalizing over network structures:

$$
\mathbb{P}(\varphi | \Theta) = \sum_{\mathbf{A} \in \mathscr{A}} \mathbb{P}(\varphi, \mathbf{A} | \Theta) = \sum_{\mathbf{A} \in \mathscr{A}} \mathbb{P}(\varphi | \mathbf{A}, \Theta) \mathbb{P}(\mathbf{A} | \Theta)
$$

\n
$$
= \sum_{\mathbf{A} \in \mathscr{A}} \mathbb{P}(\varphi | \mathbf{A}) \mathbb{P}(\mathbf{A} | \Theta) = s_2 \sum_{\mathbf{A} \in \mathscr{A}} \prod_{i=1}^{N} \prod_{j=i+1}^{N} \Phi_{ij}^{\mathbf{A}_{ij}} \mathbb{P}(\mathbf{A}_{ij} | \Theta_i, \Theta_j)
$$

\n
$$
= s_2 \prod_{i=1}^{N} \prod_{j=i+1}^{N} \sum_{\mathbf{A}_{ij} \in \{0,1\}} \Phi_{ij}^{\mathbf{A}_{ij}} \mathbb{P}(\mathbf{A}_{ij} | \Theta_i, \Theta_j)
$$

\n
$$
= s_2 \exp \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \log \left(\left\{ 1 - \mathbb{P}(\mathbf{A}_{ij} = 1 | \Theta_i, \Theta_j) \left(1 - \Phi_{ij} \right) \right\} \right) \right),
$$
\n(S7)

where the first statement makes use of the law of total probability, the second statement uses the individual-level ICAR prior in Eq. S1 and the conditional independencies in Eqs. S4 and S3, the third statement rewrites the sum over products as a product over sums, and the final statement expresses the product of terms as the exponential of the sum of their logarithms. Due to network sparsity in Eq. S5, for large *N* we can apply a first-order expansion $log(1+x) = x + \mathcal{O}(x^2) \approx x$ for $x \ll 1$, and use Eq. S2 to rewrite Eq. S7:

$$
\mathbb{P}(\varphi | \Theta) \approx s_2 \exp \left(-\frac{s_1}{2N} \sum_{i=1}^N \sum_{j=i+1}^N \left(\varphi_i - \varphi_j \right)^2 \mathbb{P} \left(\mathbf{A}_{ij} = 1 \, \middle| \, \Theta_i, \Theta_j \right) \right). \tag{S8}
$$

Hierarchical ICAR prior at the individual level We consider a statistical network model that conditions only on the regional coordinates of each individual. If H is a matrix of size $N \times m$ where each row is a one-hot encoding row vector indicating the region of residence of an individual, then assume $\Theta \triangleq H$, and let **B** be a symmetric non-negative valued matrix of size $m \times m$ such that

$$
\mathbb{P}\left(\mathbf{A}_{ij}=1\,\big|\,\mathbf{H}_{iu}=1,\mathbf{H}_{jv}=1\right)=\frac{\mathbf{B}_{uv}}{N}\tag{S9}
$$

encodes the probability of any two individuals *i* and j ($i \neq j$), respectively from regions *u* and *v*, to connect. The sparsity requirement of Eq. S5 implies that $\mathbf{B} = \mathcal{O}(1)$. Let

$$
\mathbf{p} \triangleq \frac{\mathbf{H}^T \mathbf{1}_N}{N} \tag{S10}
$$

and ϕ be vectors of size *m* encoding the population proportion residing in and mean individual random effect of every region, respectively; evidently, $\mathbf{p} = \mathcal{O}(1)$. Let

$$
\delta \triangleq \varphi - \mathbf{H}\phi \tag{S11}
$$

be a vector of size N indicating the difference of individual-level random effects φ from the regional average ϕ . Without loss of generality, we can assume that all individuals are indexed according to an arbitrary fixed ordering of regions from $\{1, 2, \dots m\}$, and let **n** be a vector of size $m+1$ encoding cumulative population size as per this ordering:

$$
\mathbf{n}_{u} \triangleq \begin{cases} 0 & \text{if } u = 1, \\ N \sum_{\nu=1}^{u-1} \mathbf{p}_{\nu} & \text{if } u \in \{2, \cdots m+1\}. \end{cases}
$$
(S12)

Using Eq. S11 the prior distribution over individual random effects in Eq. S8 can be reparameterized into a prior over regional random effects ϕ and individual-level deviations δ from them:

$$
\mathbb{P}(\phi, \delta | \mathbf{p}, \mathbf{B}) = s_2 \exp \left(-\frac{s_1}{2} \sum_{u=1}^m \sum_{v=u+1}^m (\phi_u - \phi_v)^2 \mathbf{B}_{uv} \mathbf{p}_u \mathbf{p}_v \right) \times \exp \left(-\frac{s_1}{2} \sum_{u=1}^m \frac{\mathbf{B}_{uu}}{N^2} \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=i+1}^{\mathbf{n}_{u+1}} (\delta_i - \delta_j)^2 \right) \times \exp \left(-\frac{s_1}{2} \sum_{u=1}^m \sum_{v=u+1}^m \frac{\mathbf{B}_{uv}}{N^2} \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=\mathbf{n}_v+1}^{\mathbf{n}_{v+1}} (\delta_i - \delta_j)^2 \right),
$$
\n(S13)

where the definitions in Eqs. S9, S10, and S12 were used. The statistical network model conditions only on an individual's region H, and implies that the residuals δ are uncorrelated. We further assume that conditional on **H**, individual residuals are independent and identically distributed (i.i.d.). Let σ be a non-negative valued vector of size *m* encoding the finite standard deviation in residuals for every region:

$$
\mathbb{E}\left[\delta_i\right] = 0\tag{S14a}
$$

$$
\mathbb{V}\left[\delta_i\,\middle|\,\mathbf{H}_{iu}=1\right]=\mathbb{E}\left[\delta_i^2\,\middle|\,\mathbf{H}_{iu}=1\right]=\sigma_u^2.\tag{S14b}
$$

For a region *u*, we can consider:

$$
\frac{1}{N^2} \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=i+1}^{\mathbf{n}_{u+1}} (\delta_i - \delta_j)^2 = \frac{1}{2N^2} \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} (\delta_i^2 + \delta_j^2 - 2\delta_i \delta_j)
$$
\n
$$
= \frac{1}{N^2} \left\{ (N\mathbf{p}_u - 1) \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \delta_i^2 - \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \delta_i \delta_j \right\}
$$
\n
$$
\rightarrow \mathbf{p}_u (\mathbf{p}_u - N^{-1}) \sigma_u^2 \approx \mathbf{p}_u^2 \sigma_u^2,
$$
\n(S15)

where asymptotically, for large *N*, the strong law of large numbers applies, followed by Eq. S14. Similarly for region pair *u*, *v*, we obtain:

$$
\frac{1}{N^2} \sum_{i=\mathbf{n}_u+1}^{\mathbf{n}_{u+1}} \sum_{j=\mathbf{n}_v+1}^{\mathbf{n}_{v+1}} \left(\delta_i - \delta_j\right)^2 \to \mathbf{p}_u \mathbf{p}_v \left(\sigma_u^2 + \sigma_v^2\right). \tag{S16}
$$

From Eqs. S13, S15 and S16, and using the symmetry of **B** (Eq. S6) we obtain:

$$
\mathbb{P}(\phi, \sigma | \mathbf{p}, \mathbf{B}) = s_2 \exp\left(-\frac{s_1}{4} \sum_{u=1}^m \sum_{v=1}^m \left\{ (\phi_u - \phi_v)^2 + 2\sigma_u^2 \right\} \mathbf{B}_{uv} \mathbf{p}_u \mathbf{p}_v) \right)
$$

= $s_2 \exp\left(-\frac{s_1}{2} \left\{ (\phi + \sigma)^T \text{diag}(\mathbf{D}_{\mathbf{p}} \mathbf{B} \mathbf{p}) (\phi + \sigma) - \phi^T \mathbf{D}_{\mathbf{p}} \mathbf{B} \mathbf{D}_{\mathbf{p}} \phi \right\} \right)$, (S17)

where $D_p \triangleq diag(p)$. Evidently, the prior on regional effects ϕ assume a complete covariance structure, whereas the prior on scale of residuals in each region only assume a diagonal structure, and they factorize. We drop the proportionality s_2 and use the notation from Eq. 6 to express the regional ICAR prior as:

$$
\phi \sim \mathcal{N}\left(\mathbf{0}_m, s_1^{-1} \mathbf{L}_{\mathbf{D}_{\mathbf{p}} \mathbf{B} \mathbf{D}_{\mathbf{p}}}\right) \tag{S18a}
$$

$$
\sigma_u \sim \text{HalfNormal}\left(0, \left(s_1 \left[\mathbf{D}_{\mathbf{p}} \mathbf{B} \mathbf{p}\right]_{u}\right)^{-1}\right),\tag{S18b}
$$

for any region *u*. For individual random effects, any suitable prior on the residuals δ with zero mean and bounded variance may be used. One choice is the normal distribution, which from Eq. S11 yields the prior:

$$
\varphi \sim \mathcal{N}\left(\mathbf{H}\phi, \text{diag}\left(\mathbf{H}\sigma\right)^2\right) \tag{S19}
$$

Together, Eqs. [S19](#page-7-0) and [S18](#page-7-1) provide a hierarchical prior specification for individual-level random effects, grounded in a statistical network model that conditions on individuals' region of residence.

Weighted ICAR prior at the region level For our survey sample, incorporating individuallevel random effects will imply that model parameters scale with the number of survey respondents *n*, which is large. The IRT model already assumes a logistic error distribution at the individual level (see Eq. 1b), and we do not assume additional uncertainty at the individual level due to social connections, i.e., $\sigma \to \mathbf{0}_m$ and from Eq. [S19](#page-7-0) $\varphi = \mathbf{H}\varphi$. This yields a usual ICAR prior at the region level in Eq. 5, where the weight matrix is given by $W \triangleq D_p BD_p$. In particular, the proportion of residents in each region p is given by data on population estimates at the NUTS-3 level in the UK [\[S9\]](#page-88-9) while the probability of connection B between any two NUTS-3 regions is given by Facebook's social connectedness index (SCI) [\[S4\]](#page-88-4).

Individual-level vaccine uptake model

Let n_v be the number of respondents who reported that they were invited for a COVID-19 vaccination, at the time of filling out the questionnaire (April 2021). Let y be a vector of size n_v encoding individual's vaccination status, such that $y_i = 1$ if individual *i* reports having taken at least one dose of a COVID-19 vaccine, and $y_i = 0$ otherwise. Then we model for the log-odds of being vaccinated:

$$
\mathbf{y}_{i} | \lambda_{i}^{\nu} \sim \text{Bernoulli}(\mathbf{p}_{i}), \qquad (S20a)
$$

$$
logit(\mathbf{p}_i) \triangleq log\left(\frac{\mathbf{p}_i}{1-\mathbf{p}_i}\right) = \lambda_i^{\nu}.
$$
 (S20b)

An individual's log-odds λ^{ν} of getting vaccinated are modeled analogously to an individual's latent ability to correctly identify real or fake news in the social-IRT model; see *Methods: Social-IRT model.* Let **R** and X^{ν} be matrices of size $m \times c$ and $n_{\nu} \times d$, encoding the regional and individual covariates, respectively, and G^v be a one-hot encoding matrix of size $n_v \times m$ indicating the region of residence of all individuals. Then, analogous to Eqs. 2 and 3, we can condition the log-odds on individual and regional covariates, alongside region-level random effects:

$$
\lambda^{\nu} \triangleq \mathbf{X}^{\nu} \boldsymbol{\beta}_{\mathbf{X}}^{\nu} + \mathbf{G}^{\nu} (\mathbf{R} \boldsymbol{\beta}_{\mathbf{R}}^{\nu} + \boldsymbol{\eta}^{\nu}), \qquad (S21a)
$$

$$
\eta^{\nu} \triangleq \sigma^{\nu} \left(\sqrt{1 - \rho^{\nu}} \theta^{\nu} + \sqrt{\rho^{\nu}} \phi^{\nu} \right), \tag{S21b}
$$

where β_X^{ν} and β_R^{ν} are vectors of coefficients for individual and regional covariates, of sizes *d* and *c* respectively. Fig. 2 and Table [S20](#page-83-0) show the posterior estimates for coefficients of individual and regional covariates. The vector η^{ν} of size *m* encodes region-level random effects, $\sigma^{\nu} \ge 0$ is the overall scale of random effects, θ^{ν} and ϕ^{ν} are vectors of size *m* encoding the unstructured and structured components of the random effects, respectively, and $\rho^{\nu} \in [0,1]$ indicates the proportion of variance explained by the structured component ϕ^{ν} [\[S3\]](#page-88-3). The rest of the model follows analogously from the social-IRT model: hierarchical priors are placed over coefficients of individual covariates $\beta_{\mathbf{X}}^{\nu}$, and an ICAR prior over the structured random effects ϕ^{ν} —see Eq. 5—that encodes regional similarity based on social connectivity volumes between regions. We remark that sampling settings used for the social-IRT model were re-used for this model; see *Methods: Statistical inference*.

We also note that an analogous individual-level model was used for trust in experts regarding COVID-19, with y encoding whether an individual *i* trusts $(y_i = 1)$ or does not trust $(y_i = 0)$ an expert source of COVID-19 information; see Fig. [S16.](#page-39-0) Fig. [S17](#page-40-0) and Table [S19](#page-82-0) show the corresponding posterior estimates for coefficients of individual and regional covariates.

Poststratification with income variable

The expanded covariate used for secondary analyses in GB also includes individual-level income. However, census microdata for England, Scotland and Wales does not include data on an individual's income. Data on quantiles of gross annual pay—from a sample of employees in the Annual Survey of Hours and Earnings (ASHE) for 2021 [\[S10\]](#page-88-10)—can be fit to a Burr distribution [\[S11,](#page-88-11) [S12\]](#page-88-12), typically fit to household income data, to obtain marginal distributions of annual incomes in every LAD of Great Britain. Data on the population of employeed individuals—from the Annual Population Survey (APS) for 2021 [\[S13\]](#page-88-13)—can be used to aggregate the distribution to the region level. However, poststratification requires a *joint* distribution S_X over all sociodemographic coordinates in every sub-national region [\[S14\]](#page-88-14). A synthetic joint distribution can be generated for poststratification [\[S15\]](#page-89-0) in two ways:

- Simple joint assumes no correlation between income and other socio-demographic covariates, and computes the joint distribution as a product of the marginal distribution of income and joint distribution of other covariates.
- Adjusted joint assumes access to correlation between income and other socio-demographic covariates at the national level, and computes an "adjusted" joint distribution that respects the marginals of income and joint of other covariates in every region.

We used our survey sample to generate a synthetic adjusted joint distribution of income, age, and gender [\[S15,](#page-89-0) [S16\]](#page-89-1), and used both the simple and adjusted versions for secondary analyses in Fig. [S11.](#page-34-0)

Predicting regional outcomes from MIST scores

Linear dependencies From the definition of an individual *i*'s MIST scores in Eq. 9, note that $M_v(i)$ is (linearly) dependent on $M_r(i)$, $M_f(i)$, and $M_d(i)$, $M_n(i)$ are (non-linearly) dependent on $M_r(i)$, $M_f(i)$. Since we are interested in the expectation of these scores as predictors for uptake, we consider in further detail the relationship between their expectations. Due to linearity of expectation, Eq. 10c demonstrates that $\mathbb{E}[M_r(i)]$, $\mathbb{E}[M_f(i)]$ and $\mathbb{E}[M_v(i)]$ are linearly dependent. Because poststratification performs an additional expectation, over individual socio-demographics in a given region, this linear dependence carries through and Eq. 13c shows that $\{M_r^{\mu}(u), M_f^{\mu}(u)\}$ $f(\mu)$, $M_v^{\mu}(u)$ } are linearly dependent for a region *u*. From Eqs. 9d and 9e, we also obtain a linear dependence between $\mathbb{E}[M_r(i)], \mathbb{E}[M_f(i)], \mathbb{E}[M_d(i)]$ and $\mathbb{E}[M_n(i)]$:

$$
\mathbb{E}[M_d(i)] - \mathbb{E}[M_n(i)] = \mathbb{E}[M_f(i)] - \mathbb{E}[M_r(i)]
$$

\n
$$
\implies M_d^{\mu}(u) - M_n^{\mu}(u) = M_f^{\mu}(u) - M_r^{\mu}(u).
$$
\n(S22)

From Eqs. 13c and [S22](#page-9-1) it follows that at the region level, any expected MIST score is linearly completely determined by at most three other expected MIST scores. Consequently, any model featuring a linear combination of expected MIST scores can use at most three of them as "exogenous" predictors.

Non-linear dependencies Permitting non-linear relationships reduces the set of exogenous MIST scores even further. In particular, define

$$
M_{\delta}(i) \triangleq \left| M_r(i) - M_f(i) \right| = \sqrt{\left(M_r(i) - M_f(i) \right)^2},
$$
\n(S23)

as the absolute difference of real and fake news detection ability scores for individual *i*, then the expectations of bias scores can be written from Eqs. 9d and 9e as:

$$
\mathbb{E}\left[M_d(i)\right] = \frac{\mathbb{E}\left[M_f(i)\right] - \mathbb{E}\left[M_r(i)\right] + \mathbb{E}\left[M_\delta(i)\right]}{2},\tag{S24a}
$$

$$
\mathbb{E}\left[M_n(i)\right] = \frac{\mathbb{E}\left[M_r(i)\right] - \mathbb{E}\left[M_f(i)\right] + \mathbb{E}\left[M_\delta(i)\right]}{2}.
$$
\n(S24b)

For a convex (concave) function $\check{h}(x)$ ($\hat{h}(x)$), Jensen's inequality [\[S17\]](#page-89-2) yields:

$$
\mathbb{E}\left[\check{h}(x)\right] \ge \check{h}\left(\mathbb{E}\left[x\right]\right),\tag{S25a}
$$

$$
\mathbb{E}\left[\hat{h}(x)\right] \le \hat{h}\left(\mathbb{E}\left[x\right]\right),\tag{S25b}
$$

that can be applied to the (convex) absolute and (concave) square-root functions in Eq. [S23](#page-9-2) to obtain the bounds:

$$
\left|\mathbb{E}\left[M_r(i)\right]-\mathbb{E}\left[M_f(i)\right]\right|\leq \mathbb{E}\left[M_\delta(i)\right]\leq \sqrt{\mathbb{E}\left[\left(M_r(i)-M_f(i)\right)^2\right]}.
$$
\n(S26)

The RHS of this inequality can be expanded as:

$$
\mathbb{E}\left[\left(M_r(i) - M_f(i)\right)^2\right]
$$
\n
$$
= \mathbb{E}\left[\left\{\left(M_r(i) - \mathbb{E}\left[M_r(i)\right]\right) - \left(M_f(i) - \mathbb{E}\left[M_f(i)\right]\right) + \left(\mathbb{E}\left[M_f(i)\right] - \mathbb{E}\left[M_r(i)\right]\right)\right\}^2\right] \quad (S27)
$$
\n
$$
= \mathbb{V}\left[M_r(i)\right] + \mathbb{V}\left[M_f(i)\right] + \left(\mathbb{E}\left[M_r(i)\right] - \mathbb{E}\left[M_f(i)\right]\right)^2,
$$

where we apply the modeling assumption (see *Methods: Social-IRT model*) that real and fake news detection abilities, and consequently real and fake news detection ability scores, are independent when conditioned on an individual. Since poststratification performs an additional expectation, over individual socio-demographics in a given region, the inequality in Eq. [S26](#page-10-0) carries through to the region level:

$$
\left| M_r^{\mu}(u) - M_f^{\mu}(u) \right| \le M_{\delta}^{\mu}(u) \le \sqrt{M_r^{\sigma^2}(u) + M_f^{\sigma^2}(u) + \left(M_r^{\mu}(u) - M_f^{\mu}(u) \right)^2}.
$$
 (S28)

As real and fake news detection ability scores follow Poisson binomial distributions, Eq. 12 yields a poststratified estimate of the variance in real and fake news detection ability scores:

$$
M_r^{\sigma^2}(u) = \frac{\mathbf{T}_u \left\{ \widetilde{\mathbf{P}_u} \odot \left(\mathbf{1}_t \mathbf{1}_k^T - \widetilde{\mathbf{P}_u} \right) \right\} \mathbf{1}_k}{\left[\mathbf{T} \mathbf{1}_t \right]_u},
$$
(S29a)

$$
M_f^{\sigma^2}(u) = \frac{\mathbf{T}_u \left\{ \widetilde{\mathbf{P}_u} \odot \left(\mathbf{1}_t \mathbf{1}_k^T - \widetilde{\mathbf{P}_u} \right) \right\} \mathbf{1}_k}{\left[\mathbf{T} \mathbf{1}_t \right]_u}.
$$
 (S29b)

Using Eqs. 13c, [S29,](#page-10-1) [S24](#page-9-3) and [S28,](#page-10-2) Fig. [S14](#page-37-0) shows the observed values and upper and lower bounds of M_δ^μ —the posterior mean of poststratified expectation of absolute difference between real and fake news detection ability scores. (We note that taking the posterior *mean* perform an additional expectation—over the model parameters—and therefore the arguments above follow through to the posterior means of poststratified estimates of region-level MIST scores.) Evi-dently, the upper bound (RHS of Eq. [S28\)](#page-10-2) explains almost all the variation in M_s^{μ} δ^{mu} ($R^2 = 0.997$), and from Eq. [S24](#page-9-3) we conclude that all variation in regional bias scores can be explained by the poststratified expectation and variance of real and fake news detection ability scores. Furthermore, as the function $g(x) = x(1-x)$ is concave, from Eq. S25b we obtain that the variance of a

Poisson binomial distribution is bounded from above by the variance of a binomial distribution with the same mean, yielding:

$$
M_r^{\sigma^2}(u) \le M_r^{\mu}(u) \left\{ 1 - \frac{M_r^{\mu}(u)}{k} \right\},
$$
\n(S30a)

$$
M_f^{\sigma^2}(u) \le M_f^{\mu}(u) \left\{ 1 - \frac{M_f^{\mu}(u)}{k} \right\}.
$$
 (S30b)

Fig. [S12](#page-35-0) shows how poststratified variances (LHS of Eq. [S30,](#page-11-0) given from Eq. [S29\)](#page-10-1) vary with poststratified expectations of real and fake news detection ability scores (RHS of Eq. [S30,](#page-11-0) given from Eq. 13). Combining Eqs. [S30](#page-11-0) and [S28,](#page-10-2) we get a new upper bound:

$$
M_{\delta}^{\mu}(u) \le \overline{M_{\delta}^{\mu}(u)},\tag{S31a}
$$

$$
\overline{M_{\delta}^{\mu}}(u) = \sqrt{M_{r}^{\mu}(u) + M_{f}^{\mu}(u) + \left(M_{r}^{\mu}(u) - M_{f}^{\mu}(u)\right)^{2} - \frac{M_{r}^{\mu}(u)^{2} + M_{f}^{\mu}(u)^{2}}{k}}.
$$
 (S31b)

Using Eqs. 13c, [S31](#page-11-1) and [S28,](#page-10-2) Fig. [S14](#page-37-0) shows that this upper bound $\overline{M}_{\delta}^{\mu}$ $\frac{\mu}{\delta}$ too explains almost all the variation in M_8^{μ} δ^{mu} ($R^2 = 0.997$), and from Eq. [S24](#page-9-3) we conclude that all variation in regional bias scores can be explained by the poststratified expectation of real and fake news detection ability scores:

$$
M_d^{\mu}(u) \le \overline{M_d^{\mu}}(u) = \frac{M_f^{\mu}(u) - M_r^{\mu}(u) + \overline{M_\delta^{\mu}}(u)}{2},
$$
\n(S32a)

$$
M_n^{\mu}(u) \le \overline{M_n^{\mu}}(u) = \frac{M_r^{\mu}(u) - M_f^{\mu}(u) + \overline{M_\delta^{\mu}}(u)}{2}.
$$
 (S32b)

Given the linear dependence of MIST scores in Eq. 13c, any of the following pairs of MIST scores can explain the variation in *all* expected MIST scores: $\{M_r^{\mu}, M_f^{\mu}\}$ $_{f}^{\mu}$ }, { M_{f}^{μ} $^{\mu}_{f}, M^{\mu}_{\nu}$ }, and $\{M_r^{\mu}, M_v^{\mu}\}\$. Given the quadratic dependence of bias scores on the ability scores in Eqs. [S31](#page-11-1) and [S31,](#page-11-1) we emphasize that some additional information may be required when using bias scores to estimate the ability scores. In particular, Eqs. [S32](#page-11-2) and [S31,](#page-11-1) and the linear dependencies from Eqs. 13c and [S22,](#page-9-1) yield the following:

Given
$$
\overline{M_d^{\mu}}, \overline{M_n^{\mu}} : M_v^{\mu} = k \pm \sqrt{k^2 - \left[(\overline{M_d^{\mu}} - \overline{M_n^{\mu}})^2 + 8k \overline{M_d^{\mu} M_n^{\mu}} \right]}
$$
, (S33a)

Given
$$
\overline{M_v^{\mu}}
$$
, $\overline{M_d^{\mu}}$: $M_f^{\mu} - M_r^{\mu} = 4k\overline{M_d^{\mu}} \pm \sqrt{8k(2k-1)\overline{M_d^{\mu}}^2 + M_v^{\mu}(2k-M_v^{\mu})}$, (S33b)

Given
$$
M_r^{\mu}
$$
, $\overline{M_d^{\mu}}$: $M_f^{\mu} = \frac{k(4\overline{M_d^{\mu}} + 1)}{2}$

$$
\pm \sqrt{\frac{k^2(4\overline{M_d^{\mu}} + 1)^2}{4} + M_r^{\mu}(k - M_r^{\mu}) - 4k\overline{M_d^{\mu}}(\overline{M_d^{\mu}} + M_r^{\mu})},
$$
(S33c)

where the indexing on region *u* has been suppressed for clarity. (We note that alongside the linear dependencies in Eqs. 13c and [S22,](#page-9-1) Eq. [S33](#page-11-3) can be used to solve for any MIST score given $\{\overline{M_d^{\mu}}\}$ $\{\overline{M}_d^{\overline{\mu}}, \overline{M}_n^{\overline{\mu}}\}$ or given exactly one of $\{\overline{M}_d^{\overline{\mu}}\}$ $\frac{\overline{\mu}}{d}$, $\overline{M_h^{\mu}}$ } and one of $\{M_r^{\mu}, M_f^{\mu}\}$ f^{μ}, M^{μ}_{ν} .) From Eq.

S33a, we note that all "valid" values of $\overline{M_d^{\mu}}$ $\frac{\overline{\mu}}{d}$, $\overline{M_n^{\mu}}$ yield two valid values of $M_v^{\mu} \in [0, 2k]$ along both solutions centered at *k*. That is, the bias score set $\{M_d^{\mu}\}$ $\{\overline{u}_d^{\overline{\mu}}, \overline{M_n^{\mu}}\}$ is insufficient to completely determine all MIST scores. From Eq. S33b, we note that all "valid" values of M_v^{μ} , $\overline{M_d^{\mu}}$ $\frac{\mu}{d}$ yield a valid value of $\overline{M_{\delta}^{\mu}} \ge 0 \implies M_{f}^{\mu} - M_{r}^{\mu} < 2\overline{M_{d}^{\mu}}$ $\frac{\mu}{d}$ (from Eq. S32a) only for the negative solution. That is, the ability-bias score sets $\{M_v^{\mu}, \overline{M_d^{\mu}}\}$ $\{\overline{M}_{d}^{\mu}\}$, and analogously $\{M_{\nu}^{\mu}, \overline{M_{n}^{\mu}}\}$ are sufficient to completely determine all MIST scores. Finally, from Eq. S33c, "valid" values of M_r^{μ} , $\overline{M_d^{\mu}}$ *d* yield a valid value of M_f^{μ} f ^{μ} either on just the negative solution, or on *both* positive and negative solutions. That is, the ability-bias score sets $\{M_r^{\mu}, \overline{M_d^{\mu}}\}$ $\{\overline{M}_{d}^{\mu}\},$ and analogously $\{M^{\mu}_{r}, \overline{M^{\mu}_{n}}\},$ $\{M^{\mu}_{f}\}$ $_{f}^{\mu},\overline{M_{d}^{\mu}}$ $\begin{bmatrix} \mu \\ d \end{bmatrix}$ and $\{M_f^{\mu}$ $f_{f}^{\mu}, \overline{M_n^{\mu}}\}$, are insufficient to completely determine all MIST scores. Fig. [S15](#page-38-0) shows the parameter space corresponding to the number of valid solutions in Eq. [S33.](#page-11-3) Consequently, the following pairs of MIST scores can fully explain the variation in *all* expected MIST scores— $\{M_r^{\mu},M_r^{\tilde{\mu}}\}$ $_{f}^{\mu}$, { M_{f}^{μ} $\{M_{\nu}^{\mu},M_{\nu}^{\mu}\},\,\{M_{\nu}^{\mu},M_{\nu}^{\mu}\},\,\{M_{\nu}^{\mu},M_{\nu}^{\mu}\}$ $\{\mu_d^{\mu} \}$, and $\{M_v^{\mu}, M_n^{\mu}\}$ —and are referred to as "valid" conditioning MIST sets.

Predicting outcomes using pairs of expected MIST scores Any model that considers the effects of misinformation susceptibility (i.e. MIST scores) on any outcome of interest (like vaccine uptake) must account for the described dependency structures to eliminate multicollinearity and confounding in the parameter estimates; see Table [S13](#page-76-0) for details on multicollinearity when predicting vaccine uptake rates. That is, from the set of all valid exogenous pairs of MIST score, exactly one pair must be used in predicting any given regional outcome. The choice of this pair can be made via dominance analysis [\[S18,](#page-89-3) [S19\]](#page-89-4). We choose the first predictor as the one that explains the most additional variance, when added to a subset model consisting of only the covariates. This yields fake news detection ability score M_f^{μ} f ^{μ} as the first predictor for vaccine uptake rates; see Table [S14.](#page-77-0) Now, valid exogenous pairs consisting of M_f^{μ} f_f^{μ} contain either the real news detection ability score M_r^{μ} or the veracity discernment ability score M_v^{μ} . Due to the linear dependence of the MIST scores $\{M_r^{\mu}, M_f^{\mu}\}$ f^{μ} , M^{μ}_v } (Eq. 13c), a model containing one of the MIST scores as a predictor will induce identical variances explained by the addition of a second predictor from that set. In such a scenario, we use the criterion of smallest pairwise correlation to choose the second predictor. This reduces multicollinearity and encodes the assumption of minimal common causes of the two exogenous predictors, thus improving interpretability of the regression coefficients as the effect of increasing one predictor while holding the other constant. Let $\rho_{rf} \in [-1,1]$ be the linear correlation between M_r^{μ} and M_f^{μ} f^{μ} , and $\sigma_{rf} > 0$ be the standard deviation of M_f^{μ} f_f^{μ} relative to M_r^{μ} , at the region level. Then using Eq. 13c, the linear correlation with M_v^{μ} at the region level is given by:

$$
\rho_{rv} = \frac{\sigma_{rf}\rho_{rf} + 1}{\sqrt{1 + \sigma_{rf}^2 + 2\sigma_{rf}\rho_{rf}}},
$$
\n(S34a)

$$
\rho_{fv} = \frac{\sigma_{rf} + \rho_{rf}}{\sqrt{1 + \sigma_{rf}^2 + 2\sigma_{rf}\rho_{rf}}.
$$
\n(S34b)

Fig. [S13](#page-36-0) shows how, conditioned on the first predictor chosen from $\{M_r^{\mu}, M_f^{\mu}\}$ $_{f}^{\mu}, M_{\nu}^{\mu}$ }, the parameter space for $\{\sigma_{rf}, \rho_{rf}\}$ maps to the second predictor such that the absolute value of linear correlation between the two predictors, from Eq. [S34,](#page-12-0) is minimized. Noting that $\rho_{rf} = 0.333$ and $\sigma_{rf} = 1.076$ for poststratified MIST scores in the UK, this criterion yields real news detection ability score M_r^{μ} as the second predictor for vaccine uptake rates. Consequently, we determine the effects of misinformation susceptibility on vaccine uptake rates by using regional expectation of real and fake news detection ability scores $\{M_r^{\mu}, M_f^{\mu}\}$ f_f^{μ} } as the (exogenous) predictors; see Fig. [S5](#page-28-0) for a graphical model representation.

Evaluating model predictions

Model performance Model predictions are evaluated in terms of the coefficient of determination or the R^2 value capturing the proportion of variance in true values **z** explained by the predicted values \hat{z} :

$$
R^{2}(\mathbf{z},\hat{\mathbf{z}}) \triangleq 1 - \frac{\sum_{u=1}^{m} (\mathbf{z}_{u} - \hat{\mathbf{z}}_{u})^{2}}{\sum_{u=1}^{m} (\mathbf{z}_{u} - \bar{z})^{2}},
$$
\n(S35)

where

$$
\bar{z} \triangleq \frac{\sum_{i=1}^{m} z_{u}}{m}
$$
 (S36)

indicates the mean of vector z of size *m*. Under a null model of always predicting the mean \bar{z} , clearly R^2 attains a value of 0 (no variance explained), while a larger value indicates better predictions (more variance explained), a value of 1 indicates perfect predictions (all variance explained), and a negative value indicate worse performance than the null model. For vaccine uptake models, y is the true vaccine uptake rate and μ gives the predicted (expected) vaccine uptake rate from Eqs. 15a and 17a. As Bayesian modeling is used, posterior distributions over μ yield posterior distributions over $R^2(y,\mu)$, shown in Fig. [S20.](#page-43-0) Since the outcome was standardized, we remark that $\bar{y} = 0$, yielding $R^2(y, \mu) = 1 - \frac{\theta^T \theta}{\sqrt{x}}$ $\frac{\theta^2}{y^T y}$. Model residuals were also analyzed in Fig. [S22.](#page-45-0)

Spatial autocorrelations To evaluate the existence of spatial autocorrelations based on the spatial weights matrix W, Moran's *I* statistic [\[S20\]](#page-89-5) was used:

$$
I(\mathbf{z}; \mathbf{W}) \triangleq \frac{m}{\mathbf{1}_m^T \mathbf{W} \mathbf{1}_m} \frac{\sum_{u=1}^m \sum_{v=1}^m \mathbf{W}_{uv} (\mathbf{z}_u - \overline{z}) (\mathbf{z}_v - \overline{z})}{\sum_{u=1}^m (\mathbf{z}_u - \overline{z})^2}.
$$
 (S37)

Under a null model of no spatial autocorrelations, all permutations of the vector **z** are equally likely. Let π : $\{1,2,\dotsm\}$ \rightarrow $\{1,2,\dotsm\}$ denote a (random) permutation of *m* elements, and P_{π} be the corresponding permutation matrix of size $m \times m$ such that $[P_{\pi}]_{uv} = 1$ if $\pi(v) = u$ and $[\mathbf{P}_{\pi}]_{uv} = 0$ otherwise. Noting that $\mathbf{W}_{uu} = 0$, and $\overline{\mathbf{P}_{\pi} \mathbf{z}} = \overline{\mathbf{z}}$ from Eq. [S36,](#page-13-1) the expectation of Moran's *I* statistic from Eq. [S37](#page-13-2) under the null of equally likely permutations is given by [\[S20\]](#page-89-5):

$$
\mathbb{E}\left[I(\mathbf{z};\mathbf{W})\right] = \frac{m\mathbf{1}_m^T \mathbf{W} \mathbf{1}_m}{m(m-1)\mathbf{1}_m^T \mathbf{W} \mathbf{1}_m} \mathbb{E}\left[\frac{\sum_{u=1}^m \sum_{v=1, v\neq u}^m (\left[\mathbf{P}_\pi \mathbf{z}\right]_u - \overline{z})(\left[\mathbf{P}_\pi \mathbf{z}\right]_v - \overline{z})}{\sum_{u=1}^m (\left[\mathbf{P}_\pi \mathbf{z}\right]_u - \overline{z})^2}\right]
$$
\n
$$
= \frac{1}{m-1} \mathbb{E}\left[\frac{\sum_{u=1}^m (\mathbf{z}_u - \overline{z}) \sum_{v=1}^m (\mathbf{z}_v - \overline{z}) - \sum_{u=1}^m (\mathbf{z}_u - \overline{z})^2}{\sum_{u=1}^m (\mathbf{z}_u - \overline{z})^2}\right]
$$
\n
$$
= -(m-1)^{-1},
$$
\n(S38)

where we use the definition for the mean in Eq. [S36.](#page-13-1) Thus for large *m*, under the null model of no spatial autocorrelation Moran's *I* attains a value close to 0, with larger positive (negative) values indicating larger positive (negative) spatial autocorrelations in z. Since we want to investigate spatial autocorrelations after having controlled for the fixed effects (in the spatial/structured, and non-spatial/unstructured models) and for the spatial effects (in the spatial/structured model only), Moran's *I* is computed for the residuals $\theta \triangleq v - \mu$ from Eqs. 15b and 17b. Fig. [S20](#page-43-0) show the posterior distribution over $I(\theta; \mathbf{Q})$ for the non-spatial and spatial models. Since Moran's *I* can have a large positive/negative value simply by chance, we also compute a two-sided *p*-value via a random permutation test using 1000 samples, i.e., a uniform random sample of 1000 permutations π of **z** yield the samples $I(\mathbf{P}_{\pi}\mathbf{z};\mathbf{W})$ from the null distribution of the statistic $I(z;W)$ [\[S21\]](#page-89-6). Since Bayesian modeling is used, we obtain a posterior distribution over the *p*-values, as shown in Fig. [S20.](#page-43-0) If the posterior mean of the *p*-values is small, say below $\alpha = 0.05$, then we can regard it as "significant" spatial autocorrelation at level α.

Stan programs

Code 1. Social-IRT model (United Kingdom)

```
data {
   int<lower=1> n; // number of individuals
   int<lower=1> k; // number of MIST responses per real/fake category
   int<lower=1> m; // number of subnational regions
   int<lower=1> d_reg; // number of regional covariates
    int<lower=0, upper=1> M[k, n]; // MIST real/fake responses (binary)
    // individual covariates (categoricals)
   int<lower=1> d_Age;
   int<lower=1, upper=d_Age> X_Age[n];
   int<lower=1> d_Gender;
   int<lower=1, upper=d_Gender> X_Gender[n];
   int<lower=1> d_Education;
    int<lower=1, upper=d_Education> X_Education[n];
    int<lower=1> d_Employment;
    int<lower=1, upper=d_Employment> X_Employment[n];
    int<lower=1> d_Religion;
    int<lower=1, upper=d_Religion> X_Religion[n];
   int<lower=1> d_Ethnicity;
   int<lower=1, upper=d_Ethnicity> X_Ethnicity[n];
   int<lower=1> d_Intercept;
    int<lower=1, upper=d_Intercept> X_Intercept[n];
   // other data
   matrix[m, d_reg] X_reg; // regional covariates (reals)
   matrix[m, m] L; // (scaled) Laplacian encoding regional social connectivity
    int<lower=1, upper=m> G[n]; // region of residence of individual (categoricals)
}
parameters {
   // parameters for individual ability
   vector[d_Age] beta_raw_Age;
   real<lower=0> sigma_Age;
```

```
real beta_Gender;
   vector[d_Education] beta_raw_Education;
    real<lower=0> sigma_Education;
   vector[d_Employment] beta_raw_Employment;
   real<lower=0> sigma_Employment;
   vector[d_Religion] beta_raw_Religion;
   real<lower=0> sigma_Religion;
   vector[d_Ethnicity] beta_raw_Ethnicity;
   real<lower=0> sigma_Ethnicity;
   real beta_Intercept;
    // parameters for region
   vector[d_reg] beta_region; // regional fixed effects
   vector[m] phi; // structured random effects
   vector[m] theta; // unstructured random effects
   real<lower=0, upper=1> rho; // variance explained by structured random effects
   real<lower=0> sigma_region; // overall standard deviation of random effects
    // parameters for items
   vector[k] alpha; // item difficulty
   vector<lower=0>[k] gamma; // item discrimination (constraining >0 prevents
   // multiplicative non-identifiability)
   real<lower=0> sigma_gamma; // scale of discrimination
}
transformed parameters {
   // non-centered parameters
   vector[d_Age] beta_Age;
   vector[d_Education] beta_Education;
   vector[d_Employment] beta_Employment;
   vector[d_Religion] beta_Religion;
   vector[d_Ethnicity] beta_Ethnicity;
   // regional fixed and random effects
   vector[m] eta;
   beta_Age = sigma_Age*beta_raw_Age;
   beta_Education = sigma_Education*beta_raw_Education;
   beta_Employment = sigma_Employment*beta_raw_Employment;
   beta_Religion = sigma_Religion*beta_raw_Religion;
   beta_Ethnicity = sigma_Ethnicity*beta_raw_Ethnicity;
   eta = X_reg*beta_region + sigma_region*(sqrt(1-rho)*theta + sqrt(rho)*phi));
}
model {
   vector[n] z_mu;
   // model for individual ability
   beta_raw_Age ˜ std_normal();
   sigma_Age ~ normal(0, 1);
   beta_Gender ˜ normal(0, 2);
   beta_raw_Education ˜ std_normal();
   sigma_Education ~ normal(0, 1);
   beta_raw_Employment ˜ std_normal();
   sigma_Employment ~ normal(0, 1);
   beta raw Religion ~ std normal();
   sigma_Religion ~ normal(0, 1);
   beta_raw_Ethnicity ˜ std_normal();
   sigma_Ethnicity ~ normal(0, 1);
```
}

```
beta_Intercept ~ normal(0, 2);
// model for regional effects
beta_region \tilde{ } normal(0, 1);
theta ˜ std_normal();
rho \tilde{ } beta(0.5, 0.5); // Jeffreys' prior on [0, 1]sigma_region \tilde{ } normal(0, 1);
sum(phi) ~ normal(0, 0.001*m); // soft sum-to-zero constraint
target += -0.5*(phi'*L*phi); // add ICAR log-probability to the model
// model for items
alpha \tilde{ } std_normal(); // fix the scale of difficulties (for identifiability)
gamma ˜ lognormal(0, sigma_gamma);
sigma_gamma ~ normal(0, 1);
z_mu = beta_Age[X_Age] + [beta_Gender, -beta_Gender][X_Gender]' +
       beta_Education[X_Education] + beta_Employment[X_Employment] +
       beta_Religion[X_Religion] + beta_Ethnicity[X_Ethnicity] + beta_Intercept +
       eta[G]; // mean individual ability to detect real/fake news
for (i \text{ in } 1:k)M[i] \sim bernoulli_logit(gamma[i] \star (z_mu-alpha[i]));
```
Code 2. Vaccine uptake model (non-spatial)

```
data {
    int<lower=1> n; // number of subnational regions
    // regional covariates and/or MIST scores
    int<lower=1> d_M;
    matrix[n, d_M] M;
    int<lower=1> d_X;
    matrix[n, d_X] X;
    vector[n] Y; // regional vaccine uptake rate
}
parameters {
    // fixed effects for covariates and/or MIST scores
    vector[d_M] beta_M;
    vector[d_X] beta_X;
    real<lower=0> sigma; // standard deviation of random effects
}
transformed parameters {
    // regional fixed and random effects
    vector[n] mu = M*beta_M + X*beta_X; // fixed effects
    vector[n] theta = Y - mu; // random effects
}
model {
    // model for regional effects
    beta_M \sim normal(0, 1);
    betaX \sim normal(0, 1);
    sigma \tilde{ } normal(0, 1);
    theta \tilde{ } normal(0, sigma);
}
generated quantities {
```

```
real Y_{pred}[n] = normal_{rng}(mu, sigma);
```
Code 3. Vaccine uptake model (spatial)

```
data {
    int<lower=1> n; // number of subnational regions
    int<lower=1> m; // number of edges between subnational regions
    int<lower=1> c; // number of connected components of regional adjacency network
    // regional covariates and/or MIST scores
    int<lower=1> d_M;
    matrix[n, d_M] M;
    int<lower=1> d_X;
    matrix[n, d_X] X;
    vector[n] Y; // regional vaccine uptake rate
    int<lower=1, upper=n> node1[m]; // encoding source nodes of edges
    int<lower=1, upper=n> node2[m]; // encoding target nodes of edges
    int<lower=1, upper=n> comp[n]; // node indices ordered as per component memberships
    int<lower=1, upper=n> compsize[c]; // sizes of connected components
    real<lower=0> scale; // scale for unit marginal variances of spatial random effects
}
parameters {
    // fixed effects for covariates and/or MIST scores
    vector[d_M] beta_M;
    vector[d_X] beta_X;
    vector[n] phi; // spatial random effects
    real<lower=0, upper=1> rho; // variance explained by spatial random effects
    real<lower=0> sigma; // overall standard deviation of random effects
}
transformed parameters {
    // regional fixed and random effects
    real scale_theta = sigma*sqrt(1-rho); // scale of non-spatial random effects
    real scale_phi = sigma*sqrt(rho/scale); // scale of spatial random effects
    vector[n] mu = M*beta_M + X*beta_X + scale_phi*phi; // fixed and spatial random
    // effects
    vector[n] theta = Y - mu; // non-spatial random effects
}
model {
    int pos = 1; // dummy index to sum over connected components
    // model for regional effects
    beta_M \sim normal(0, 1);
    betaX \sim normal(0, 1);
    rho \tilde{ } beta(4, 4); // regularizing prior on [0, 1]
    sigma \tilde{ } normal(0, 1);
    theta ˜ normal(0, scale_theta);
    // soft sum-to-zero constraint on phi in every connected component
    for (i \text{ in } 1:c) {
        sum(phi[segment(comp, pos, compsize[i])]) ~ normal(0, 0.001*compsize[i]);
        pos = pos + compsize[i];
    }
```
}

```
target += -0.5*dot_self(phi[node1] - phi[node2]); // add ICAR log-probability
}
generated quantities {
   real Y_pred[n] = normal_rng(mu, scale_theta);
}
```
Code 4. Individual-level vaccine uptake model

```
data {
   int<lower=1> n; // number of individuals
   int<lower=1> m; // number of subnational regions
    int<lower=1> d_reg; // number of regional covariates
    int<lower=0, upper=1> V[n]; // vaccination outcome (binary)
    // individual covariates (categoricals) and/or MIST scores
    int<lower=1> d_Age;
    int<lower=1, upper=d_Age> X_Age[n];
    int<lower=1> d_Gender;
    int<lower=1, upper=d_Gender> X_Gender[n];
    int<lower=1> d_Education;
    int<lower=1, upper=d_Education> X_Education[n];
    int<lower=1> d_Employment;
    int<lower=1, upper=d_Employment> X_Employment[n];
    int<lower=1> d_Religion;
    int<lower=1, upper=d_Religion> X_Religion[n];
    int<lower=1> d_Ethnicity;
    int<lower=1, upper=d_Ethnicity> X_Ethnicity[n];
    int<lower=1> d_Income;
    int<lower=1, upper=d_Income> X_Income[n];
    int<lower=1> d_Intercept;
    int<lower=1, upper=d_Intercept> X_Intercept[n];
    int<lower=1> d_M;
   matrix[n, d_M] M;
    // regional covariates
   matrix[m, d_reg] X_reg; // regional covariates (reals)
   matrix[m, m] L; // (scaled) Laplacian encoding regional social connectivity
    int<lower=1, upper=m> G[n]; // region of residence of individual (categoricals)
}
parameters {
    // parameters for individual log-odds to be vaccinated
   vector[d_Age] beta_raw_Age;
    real<lower=0> sigma_Age;
    real beta_Gender;
    vector[d_Education] beta_raw_Education;
    real<lower=0> sigma_Education;
   vector[d_Employment] beta_raw_Employment;
    real<lower=0> sigma_Employment;
   vector[d_Religion] beta_raw_Religion;
    real<lower=0> sigma_Religion;
    vector[d_Ethnicity] beta_raw_Ethnicity;
    real<lower=0> sigma_Ethnicity;
   vector[d_Income] beta_raw_Income;
    real<lower=0> sigma_Income;
```

```
real beta_Intercept;
    vector[d_M] beta_M;
    // parameters for region
    vector[d_reg] beta_region; // regional fixed effects
    vector[m] phi; // spatial random effects
    vector[m] theta; // non-spatial random effects
    real<lower=0, upper=1> rho; // variance explained by spatial random effects
    real<lower=0> sigma_region; // overall standard deviation of random effects
}
transformed parameters {
    // non-centered parameters
    vector[d_Age] beta_Age;
    vector[d_Education] beta_Education;
    vector[d_Employment] beta_Employment;
    vector[d_Religion] beta_Religion;
    vector[d_Ethnicity] beta_Ethnicity;
    vector[d_Income] beta_Income;
    // regional fixed and random effects
    vector[m] eta;
    beta_Age = sigma_Age*beta_raw_Age;
    beta_Education = sigma_Education*beta_raw_Education;
    beta_Employment = sigma_Employment*beta_raw_Employment;
    beta_Religion = sigma_Religion*beta_raw_Religion;
    beta_Ethnicity = sigma_Ethnicity*beta_raw_Ethnicity;
    beta_Income = sigma_Income*beta_raw_Income;
    eta = X_reg * beta\_region + sigma\_region * (sqrt(1-rho) *theta + sqrt(rho) *phi);
}
model {
    vector[n] v_mu;
    // model for individual log-odds
    beta_raw_Age ˜ std_normal();
    sigma_Age \tilde{ } normal(0, 1);
    beta_Gender ˜ normal(0, 2);
    beta_raw_Education ˜ std_normal();
    sigma_Education ~ normal(0, 1);
    beta_raw_Employment ˜ std_normal();
    sigma_Employment ~ normal(0, 1);
    beta_raw_Religion ˜ std_normal();
    sigma_Religion ~ normal(0, 1);
    beta_raw_Ethnicity ˜ std_normal();
    sigma_Ethnicity ~ normal(0, 1);
    beta_raw_Income ˜ std_normal();
    sigma_Income \tilde{ } normal(0, 1);
    beta_Intercept ˜ normal(0, 2);
    betaM \sim normal(0, 1);
    // model for regional effects
   beta_region \sim normal(0, 1);
    theta ˜ std_normal();
    rho \tilde{ } beta(0.5, 0.5); // Jeffreys' prior on [0, 1]
    sigma_region \tilde{ } normal(0, 1);
    sum(phi) ~ normal(0, 0.001*m); // soft sum-to-zero constraint
    target += -0.5*(phi'*L*phi); // add ICAR log-probability to the model
```

```
v_mu = beta_Age[X_Age] + [beta_Gender, -beta_Gender][X_Gender]' +
       beta Education[X Education] + beta Employment[X Employment] +
       beta_Religion[X_Religion] + beta_Ethnicity[X_Ethnicity] +
       beta Income[X Income] + beta Intercept + M*beta M + eta[G];
       // mean individual log-odds to be vaccinated
V ~ bernoulli logit(v mu);
```
Questionnaire

}

Socio-demographics

We will begin by asking you some questions about yourself.

DEMOPC Enter first half of your postcode. *(Outward postcode (OPC) is authenticated against a list of valid OPCs.)*

DEMAGENUM How old are you? *(Numeric values from 18 to 100.)*

DEMSEX I am

- 1. Male
- 2. Female
- 3. Other
- DEMEDU What is the highest level of education you have completed? (Select the response that best applies)
	- 1. No academic qualifications
	- 2. 0-4 GCSE, O-levels, or equivalents
	- 3. 5+ GCSE, O-levels, 1 A level, or equivalents
	- 4. Apprenticeship
	- 5. 2+ A levels or equivalents
	- 6. Undergraduate or postgraduate degree, or other professional qualification
	- 7. Other (e.g. vocational, foreign qualifications)
	- 8. Do not know
	- 9. Do not wish to answer

DEMWRK Which of the following best describes your work status 6 months ago?

- 1. Working full-time (including self-employed)
- 2. Working part-time (including self-employed)
- 3. Unemployed
- 4. Student
- 5. Looking after the home
- 6. Retired
- 7. Unable to work (including, for example, a short- or long-term disability)
- 8. Do not wish to answer

DEMREL Do you consider yourself

- 1. Christian
- 2. Hindu
- 3. Muslim
- 4. Jewish
- 5. Buddhist
- 6. Atheist or agnostic
- 7. Other
- 8. Do not wish to answer

DEMETH Which best describes your ethnicity (select the response that best applies)

- 1. White: English/Welsh/Scottish/Northern Irish/British
- 2. White: Irish
- 3. White: Other white background
- 4. White and Black Caribbean
- 5. White and Black African
- 6. White and Asian or White and Asian British
- 7. Black, African, Caribbean or Black British
- 8. Asian or Asian British: Indian
- 9. Asian or Asian British: Pakistani
- 10. Asian or Asian British: Chinese
- 11. Asian or Asian British: Other
- 12. Gypsy or Irish traveller
- 13. Other
- 14. Do not wish to answer
- 15. Roma

DEMINC What is your total household income in GBP (E) from all sources before tax?

- 1. Under £15,000
- 2. £15,000 to £24,999
- 3. £25,000 to £34,999
- 4. £35,000 to £44,999
- 5. £45,000 to £54,999
- 6. £55,000 to £64,999
- 7. £65,000 to £99,999
- 8. Over £100,000
- 9. Do not wish to answer

COVID-19 vaccination status

We will now ask you some questions about coronavirus (COVID-19) and new COVID-19 vaccines.

- 1. Yes
- 2. No
- 3. Do not know

COV DOSE *(If COV INV=1)* Have you had at least one dose of a coronavirus (COVID-19) vaccine?

- 1. Yes, I have had one dose
- 2. Yes, I have had both doses
- 3. No

COV INFO What are your main sources for information about coronavirus (COVID-19) and or a coronavirus vaccine? (select all that apply)

- 1. National television
- 2. Satellite / international television channels
- 3. Radio
- 4. Newspapers
- 5. Social media (e.g Facebook, Twitter, etc)
- 6. National public health authorities (e.g. the NHS or Public Health England / Wales)
- 7. Healthcare workers (e.g. doctors, nurses, etc)
- 8. International health authorities (e.g. The World Health Organization)
- 9. Government websites
- 10. The internet or search engines (e.g. Google)
- 11. Family and friends
- 12. Work, school, or college
- 13. Other (please specify)
- 14. Do not know

Misinformation susceptibility test (MIST)

We would now like to ask you some questions on perceptions towards news headlines. Please categorise the following news headlines as either "fake news" or "real news". Some items may look credible or obviously false at first sight, but may actually fall in the opposite category. However, for each news headline, only one category is correct. *(The following items are displayed in a random ordering, and for each item two options are presented: "Fake" and "Real". Note that the first (last) 10 headlines are actually fake (real).)*

- *MIST1* Government Officials Have Manipulated Stock Prices to Hide Scandals
- *MIST2* The Corporate Media Is Controlled by the Military-industrial Complex: The Major Oil Companies Own the Media and Control Their Agenda
- *MIST3* New Study: Left-Wingers Are More Likely to Lie to Get a Higher Salary
- *MIST4* The Government Is Manipulating the Public's Perception of Genetic Engineering in Order to Make People More Accepting of Such Techniques
- *MIST5* Left-Wing Extremism Causes 'More Damage' to World Than Terrorism, Says UN Report
- *MIST6* Certain Vaccines Are Loaded with Dangerous Chemicals and Toxins
- *MIST7* New Study: Clear Relationship Between Eye Color and Intelligence
- *MIST8* The Government Is Knowingly Spreading Disease Through the Airwaves and Food Supply
- *MIST9* Ebola Virus 'Caused by US Nuclear Weapons Testing', New Study Says
- *MIST10* Government Officials Have Illegally Manipulated the Weather to Cause Devastating Storms
- *MIST11* Attitudes Toward EU Are Largely Positive, Both Within Europe and Outside It
- *MIST12* One-in-Three Worldwide Lack Confidence in NGOs
- *MIST13* Reflecting a Demographic Shift, 109 US Counties Have Become Majority Nonwhite Since 2000
- *MIST14* International Relations Experts and US Public Agree: America Is Less Respected Globally
- *MIST15* Hyatt Will Remove Small Bottles from Hotel Bathrooms by 2021
- *MIST16* Morocco's King Appoints Committee Chief to Fight Poverty and Inequality
- *MIST17* Republicans Divided in Views of Trump's Conduct, Democrats Are Broadly Critical
- *MIST18* Democrats More Supportive than Republicans of Federal Spending for Scientific Research
- *MIST19* Global Warming Age Gap: Younger Americans Most Worried
- *MIST20* US Support for Legal Marijuana Steady in Past Year

Figs. S1 to S23

Fig. S1. Overview of the study and statistical modeling. A large-sample $(n = 16,477)$ nationally representative survey was used to test whether misinformation susceptibility, as measured by the psychometrically validated misinformation susceptibility test (MIST [\[S22,](#page-89-7) [S23\]](#page-89-8)), informs actual COVID-19 vaccine uptake at the region level of grouped NUTS-3 regions in the United Kingdom (UK). Given the survey responses, "social" item response theory (IRT) models inferred latent abilities of individuals to correctly identify (1) real and (2) fake news items, conditioned on a large set of individual and regional covariates (Fig. 3), and the online social connectivity volumes between regions. These two abilities define five "MIST scores" that measure different dimensions of misinformation susceptibility; census microdata were used to aggregate scores to the region level via poststratification [\[S14\]](#page-88-14) (Fig. 4). Regional MIST scores were then used as predictors for uptake rates of the second dose of a COVID-19 vaccine in England and Scotland to test the primary hypotheses (Fig. 1) by considering the highest posterior density intervals (HPDI) of the coefficients against a preregistered region of practical equivalence (ROPE) to null effects [\[S24\]](#page-89-9). While this study was not designed to measure causal effects, a large set of potential regionlevel confounders were controlled for, including spatial autocorrelations, that ensures the robustness of results.

Fig. S2. Graphical representation for the social-IRT models inferring individuals' abilities to detect real and fake news. Observed variables are shaded gray, latent variables are colored white, and hyperparameters are excluded for clarity. Variables specific to fake news detection appear underlined. The 2-parameter item response theory models encode for the latent item parameters of "difficulty" α and "discrimination" γ , and the latent ability of individuals λ to correctly identify a real/fake news headline Y/Y : see Eq. 1. The latent ability is a function of individual covariates X, regional covariates R and random effects η corresponding to the individual's region of residence G: see Eq. 2. Regional random effects η are decomposed into an unstructured component θ , and a structured component ϕ that accounts for regional covariance as per inter-regional "connectivity strengths" W: see Eqs. 3 and 5. The parameter ρ encodes the contribution of structured random effects, while σ encodes overall scale of random effects. This graphical model encodes the conditional independence of Y and \underline{Y} given the individuals' abilities λ and λ , allowing for two separate models to be inferred for real and fake news detection abilities respectively.

Fig. S3. Graphical representation for individual random effects under a statistical network model. Individual-level random effects φ depend on whether there is a direct connection between two individuals in the social network, as encoded in the network's adjacency matrix A: see Eq. S1. In turn, the existence of a connection between any two individuals depends entirely on there location in some *d*-dimensional space Θ: see Eq. S5. This graphical model encodes the conditional independencies in Eqs. S3 and S4.

Fig. S4. Graphical representation for poststratifying MIST scores to the region level. The expectation of the misinformation susceptibility test (MIST) scores of an individual is given by the probabilities P/P of correctly identifying real/fake news headlines, conditioned on the individual's sociodemographics and region of residence: see Eq. 10. The counts T of people in a given region with a given socio-demographic coordinate can be used to poststratify the expectation of MIST scores to the region level: see Eq. 13.

Fig. S5. Graphical representation for vaccine uptake models. Observed variables are shaded gray, latent variables are colored white, and hyperparameters are excluded for clarity. (Top left) The non-spatial model considers every region's vaccine uptake rate y as an independent outcome conditioned on the MIST scores M, while controlling for regional covariates R that are expected the confound the effect of M on y: see Eq. 15. (Top right) The spatial model allows for structured random effects ϕ conditioned on regional adjacency matrix Q, thus controlling for regional adjacency that is expected to confound the effect of M on y: see Eq. 17. (Bottom) Analyzing the dependency structure of MIST scores, the dominance of MIST scores in determining vaccine uptake, and the multicollinearity between pairs of MIST scores, reveals that the regional expectation of real and fake news detection ability scores (M_r^{μ}, M_f^{μ}) $f(t)$ completely determine the regional expectations of remaining MIST scores: see *[Supplementary methods:](#page-4-0) [Predicting](#page-9-0) [regional outcomes from MIST scores](#page-9-0)*. Thus, models determining the effect of regional misinformation susceptibility on regional vaccine uptake use M_r^{μ} , M_f^{μ} f_{f}^{μ} as the (exogenous) MIST predictors—eliminating confounding due to underlying common causes of susceptibility *U*.

(a) Small β_M^z consistent with no confounding

(b) Large β_M^z consistent with confounding

(c) Large β_M^z consistent with no confounding

Fig. S6. Causal graphical model for the placebo outcome analysis. Observed variables misinformation susceptibility *M*, vaccine uptake *y*, and placebo outcome *z*—are shaded gray while unobserved (latent) variables—willingness to trust expert advice *U*—are colored white. When considering the effect of *M* on *y*, say we have an unobserved suspected confounder *U*, i.e. *U* is known to affect *y* but we are unsure about whether it influences *M* and thus whether it confounds the effect of *M* on *y*. Then the confounding can be ruled out via placebo outcome analysis [\[S25\]](#page-89-10) by considering a "placebo" outcome *z* which is directly influenced by *U* but not by *M*, i.e. model (a) holds but model (c) does not hold. For instance, levels of obesity *z* are influenced by trust in doctors *U* [\[S26\]](#page-89-11) but likely not by misinformation *M*, making it a "good" placebo outcome candidate. Then, the coefficient β_M^z when predicting *z* using *M* being "small" is more consistent with model (a)—where *U* does not directly influence *M* and therefore does not confound the effect of *M* on *y*—than with model (b)—where *U* directly influences *M* and therefore confounds the effect of *M* on *y* and consequently needs to be controlled for causal identification of the effect of *M* on *y*.

Fig. S7. Random effects of the social-IRT models based on social connectivity volumes between regions, for modeling real (**a**-**c**) and fake (**d**-**f**) news detection abilities, mapped across the United Kingdom (UK). Panels **a**, **d** indicate the structured effect, **b**, **e** indicate the unstructured effects, and **c**, **f** indicate the total random effect of the models. Colors indicate posterior means of the random effects.

Fig. S8. Being older, more educated, atheist, of white ethnicity, and earning higher income, are individual characteristics associated with improved abilities to detect real and fake news headlines in Great Britain. Those who are students or retired have higher real and fake news detection abilities, while those who are Asian or have a mixed ethnicity have lower real and fake news detection abilities. Being male is associated with a higher real news detection ability, while being Black or Muslim is associated with a lower fake news detection ability. At the region level, residing in a region with a larger proportion of those who voted to "remain" in the EU referendum is strongly indicative of a higher real news detection ability, whereas those in densely populated regions have a lower fake news detection ability. Panels correspond to different individual covariates, except for the last panel "**Region**" that corresponds to regional covariates. "Null" corresponds to undisclosed sociodemographic identity; see Table [S2](#page-48-0) for full variable recodes. Markers indicate posterior means while bars indicate 95% HPDIs. For each individual covariate, the reference group is indicated by a bounding box and markers at 0 and bars of 0 length. See Table [S9](#page-64-0) for full posterior values.

Fig. S9. Misinformation susceptibility test (MIST) scores mapped across the United Kingdom (UK). Panels **a**-**e** indicate poststratified regional estimates of expected ability scores of real news detection (M_r^{μ}, \mathbf{a}) , fake news detection (M_f^{μ}) $f(t)$, **b**), veracity discernment (M_v^{μ}, \mathbf{c}) , and expected bias scores of distrust (M_d^{μ}) d_i , **d**), and naivety (M_n^{μ}, \mathbf{e}) . Colors indicate posterior means of the poststratified estimate of expected MIST scores in every region, with blue (red) indicating higher ability (bias) scores and thus lower (higher) misinformation susceptibility. See Table [S11](#page-66-0) for full posterior values.

Fig. S10. NUTS-1 geographies (left) and NUTS-3 regions (right) of the United Kingdom (UK) [\[S27\]](#page-89-12). The UK is divided into 12 geographies at the first administrative level (NUTS-1), subdivided further into 179 regions at the third level (NUTS-3) [\[S28\]](#page-90-0). Of these 179 regions, some were combined to yield 149 "grouped" NUTS-3 regions (see Table [S6\)](#page-58-0) to align with census microdata geographies [\[S29–](#page-90-1)[S31\]](#page-90-2) to facilitate poststratification.See Tables [S4](#page-50-0) & [S6](#page-58-0) for names of NUTS-1 geographies and NUTS-3 regions.

Fig. S11. Regional MIST scores are robust to different inclusion criteria for the United Kingdom, and to using an expanded set of covariates for Great Britain (GB). *x*-axis indicates posterior estimates of the poststratified expected MIST scores in the primary model, while the *y*-axis indicates raw survey estimates (**a**) and poststratified estimates from robustness checks (**b**, **c**) and secondary models (**d**, **e**). Columns refer to different MIST scores: real news detection M_r^{μ} , fake news detection M_f^{μ} f^{μ} , veracity discernment M_{ν}^{μ} , distrust M_{d}^{μ} μ_d^{μ} , and naivety M_n^{μ} . Row (**a**) indicates raw estimates of the mean MIST scores from the survey sample: markers indicate the sample mean on the *y*-axis while bars along the *y*-axis indicate 95% Bayesian confidence intervals of the mean [\[S32,](#page-90-3) [S33\]](#page-90-4). Rows (**b**, **c**) indicate results when social-IRT models are inferred with different threshold *w* indicating the probability of an individual with a given outward postcode to map to a given grouped NUTS-3 region, since some outward postcodes may match to multiple grouped NUTS-3 regions. The primary model uses $w = 0.5$ (50 samples removed), whereas robustness checks consider more strict thresholds of $w = 0.7$ (b) and $w = 0.9$ (c), with 765 and 2343 samples removed in total, respectively. Rows (**d**, **e**) indicate the results of secondary models for England, Scotland and Wales using the expanded covariate set. In particular, poststratification along an individual's income is performed either by assuming a factorizable joint distribution (**d**) or assuming an adjusted joint distribution (**e**) (see *[Supplementary methods:](#page-4-0) [Poststratification with income variable](#page-8-1)*). For rows (**b-e**) markers indicate posterior means while bars indicate 95% HPDIs. ρ indicates the linear correlation coefficient between markers' values on the *x*- and *y*-axes.

Fig. S12. Bivariate relationships of regional MIST statistics in the United Kingdom. Markers indicate posterior means of poststratified estimates of pairs of statistics for every region, with contours indicating bivariate joint densities and bar plots indicating univariate marginal distributions. Poststratified expectation (variance) of MIST scores is computed from Eq. 13 (Eq. [S29\)](#page-10-1). The last row and column correspond to the expected fake news detection ability score computed after dropping the vaccine-specific fake news item (see *[MIST6](#page-23-1)* in *[Questionnaire](#page-20-0)*) from the summation in Eq. 13b; its near perfect correlation with the original expected fake news detection ability score indicates robustness of the MIST to individual items. See Table [S13](#page-76-0) for more details on multicollinearity when predicting vaccine uptake rates from MIST scores.

 M_2 such that absolute correlation $|\rho_{\mathsf{M}_1\mathsf{M}_2}|$ is smallest for given M_1

 Log_2 of standard deviation of M_f^{μ} relative to M_f^{μ}

Fig. S13. When using the regional expected fake news detection ability score as the first predictor in regressing for regional outcomes, using the regional expected real news detection ability score as the second predictor minimizes collinearity when predicting regional outcomes. *x*-axis indicates logarithm of the standard deviation σ_{rf} of M_f^{μ} f_f^{μ} relative to M_r^{μ} , and *y*-axis indicates the linear correlation ρ_{rf} between M_r^{μ}, M_f^{μ} f_t^{μ} . (Top) Parameter space identifying the second predictor having the smallest absolute value of linear correlation with the given first predictor: see Eq. [S34.](#page-12-0) (Bottom) Smallest possible absolute value of linear correlation between the second predictor and the given first predictor. Marker \times indicates the location $\sigma_{rf} = 1.076$, $\rho_{rf} = 0.333$ corresponding to the posterior means of the poststratified expectation of real and fake news detection ability scores in the United Kingdom.

Fig. S14. Regional expectation of real and fake news detection ability scores explain all variation in regional expectation of distrust and naivety bias scores. *x*-axis indicates the observed posterior means of the poststratified expectation of $M_{\delta}^{\mu} = \left|M_f^{\mu} - M_r^{\mu}\right|$, and *y*-axis indicates the bounds on M_{δ}^{μ} . Eq. [S28](#page-10-0) (lower and upper bounds) and from Eq. [S31](#page-11-0) (upper bound using only means/expectations). The $\frac{\mu}{\delta}$ from upper bound using only the expectations explains all the variation in M_{δ}^{μ} $\frac{\mu}{\delta}$ ($R^2 = 0.997$), i.e. from Eq. [S32](#page-11-1) the regional expectation of real and fake news detection ability scores explain all variation in the regional expectation of bias scores.

Fig. S15. MIST bias scores are insufficient as exogenous MIST predictors when predicting regional outcomes. For each given pair of regional MIST scores, each panel shows the parameter space that yields one or more valid solutions for the remaining MIST scores, using Eq. [S33](#page-11-2) with $k = 10$ items. Regional distrust bias score M_d^{μ} d/d completely determines the remaining MIST scores only alongside the veracity discernment ability score M_v^{μ} .

Fig. S16. Sources of information regarding the COVID-19 pandemic naturally cluster into expert and non-expert sources based on whether they are trusted. Binary responses indicating which sources of information regarding the COVID-19 pandemic were trusted by survey respondents (see *[Question](#page-20-0)[naire](#page-20-0)*) were used to compute a "distance" (color) between every source pair *a*, *b*—given by $\frac{1-\phi_{ab}}{2}$ where $\varphi_{ab} \in [-1,1]$ indicates the phi-coefficient, a balanced measure of similarity between binary responses [\[S34\]](#page-90-0) equivalent to the standard Pearson's correlation coefficient for binary responses—and clustered using agglomerative clustering [\[S35\]](#page-90-1) with the average linkage criterion [\[S36\]](#page-90-2) as implemented in SciPy [\[S33\]](#page-90-3). Consequently, the sources of information can be naturally bi-partitioned into "expert" sources government websites, public health authorities and healthcare workers (see black square)—and "nonexpert" sources—the rest.

Fig. S17. Individual-level regression models for self-reported trust in an expert source of COVID-19 information. Markers indicate posterior means while bars indicate 95% HPDIs of the coefficients of individual socio-demographics and regional fixed effects ("**Region**"), while including random effects structured by social connectivity volumes. For each individual covariate, the reference group is indicated by a bounding box and markers at 0 and bars of 0 length. "Null" corresponds to undisclosed sociodemographic identity; see Table [S2](#page-48-0) for full variable recodes. The dotted line indicates the reference value of 0. See Table [S19](#page-82-0) for full posterior values.

Regional vaccine uptake (Dose 2; %)

Fig. S18. Bivariate relationship of regional vaccine uptake with regional poststratified model estimates of various key statistics. *x*-axis indicates observed regional COVID-19 vaccine uptake rates across 129 regions of England and Scotland [\[S37\]](#page-90-4), while *y*-axis indicates regional poststratified model estimates of key statistics: real (M_r^{μ}) ; **a**) and fake (M_f^{μ}) f^{μ} ; **b**) news detection ability scores, the probability of being vaccinated as derived from self-reported vaccination status (**c**; see *[Questionnaire](#page-20-0)*), and the probability of trusting an expert source of COVID-19 information (**d**; see Fig. [S16\)](#page-39-0). Pearson's correlation coefficient is indicated in the bottom-right of each sub-panel. Evidently, there is a strong positive relationship between self-reported and actual uptake rates (**b**) and fake news detection ability scores and actual uptake rates (**c**).

Fig. S19. Posterior estimates of standardized coefficients in ^a secondary set of regression models for COVID-19 vaccine's second dose (while controlling for trust in experts;**^a**) and first dose regional uptake rates in England and Scotland (at two timepoints;**b**, **^c**), and two placebo**outcomes in England (d, e).** Markers indicate posterior means while bars indicate 95% HPDIs of the standardized coefficients of real (M_r^{μ}) and fake (M_f^{μ}) news detection ability scores, both before (blue) and after (red) controlling for spatial effects of regional adjacency, with the green shaded region indicating the preregistered region of practical equivalence (ROPE) to null effects of ([−]⁰.05,0.05). The models control for ^a set of regional covariates,and dotted line indicates the reference value of 0. See Tables [S15](#page-78-0) and [S17](#page-80-0) for full posterior values.

Fig. S20. Posterior distributions of R^2 of vaccine uptake and Moran's I statistic for different vaccine uptake models. Primary analyses considered second dose uptake as of 1 October 2021 (**a**, **b**), while secondary analyses considered first dose uptake as of 1 October 2021 (**c**, **d**) and 1 July 2021 (**e**, **f**), both before (**a**, **c**, **e**) and after (**b**, **d**, **f**) controlling for spatial effects of regional adjacency. All models include regional real and fake news detection ability scores as predictors, while controlling for a set of regional covariates. For the two leftmost columns, markers indicate posterior means while bars indicate 95% HPDIs, while the other columns display entire posterior distributions. Posterior means of the *p*-value (rightmost column) for Moran's *I* are less than 0.05 for non-spatial models, indicating "significant" spatial structure in the residuals of model predictions at level $\alpha = 0.05$. Mean of the posterior of R^2 values is smaller throughout the non-spatial models when compared to spatial models, suggesting that the spatial structure explains a significant amount of variance in uptake. For a residual analysis see Fig. [S22.](#page-45-0)

Fig. S21. Posterior distributions of R^2 of vaccine uptake (a, b) and other placebo outcomes (c-f), and Moran's *I* statistic for different regional regression models. Additional analyses checked for explicit confounding by considering the outcome of second dose uptake as of 1 October 2021 while controlling for trust in expert sources of COVID-19 information (**a**, **b**), while placebo analyses considered various placebo outcomes that are expected to be influenced by trust in expert authority or willingness to follow expert advice but *not* by misinformation exposure—namely the proportion of obese or overweight adult population (**c**, **d**) and the proportion of physically active adult population (**e**, **f**)—both before (**a**, **c**, **e**) and after (**b**, **d**, **f**) controlling for spatial effects of regional adjacency. All models include regional real and fake news detection ability scores as predictors, while controlling for a set of regional covariates. For the two leftmost columns, markers indicate posterior means while bars indicate 95% HPDIs, while the other columns display entire posterior distributions. Posterior means of the *p*-value (rightmost column) for Moran's *I* are less than 0.05 for non-spatial models, indicating "significant" spatial structure in the residuals of model predictions at level α = 0.05. Mean of the posterior of R^2 values is smaller throughout the non-spatial models when compared to spatial models, suggesting that the spatial structure explains a significant amount of variance in outcomes considered. For a residual analysis see Fig. [S23.](#page-46-0)

Regional vaccine uptake: predicted mean

Fig. S22. Posterior means of standardised residuals of vaccine uptake for different vaccine uptake models. Primary analyses considered second dose uptake as of 1 October 2021 (**a**, **b**), while secondary analyses considered first dose uptake as of 1 October 2021 (**c**, **d**) and 1 July 2021 (**e**, **f**), both before (**a**, **c**, **e**) and after (**b**, **d**, **f**) controlling for spatial effects of regional adjacency. All models include regional real and fake news detection ability scores as predictors, while controlling for a set of regional covariates. Main plots show the posterior mean of the standardised residuals (on the *y*-axis) against the posterior mean of the outcome (on the *x*-axis)—we observe no apparent trend in the residuals suggesting homoscedasticity. Inset plots show the quantile-quantile plot for the standard normal distribution on the *x*-axis and for the (posterior mean of) standardised residuals on the *y*-axis, with the reference line in red—we observe a good fit suggesting that residuals are normally distributed.

Regional outcome: predicted mean

Fig. S23. Posterior means of standardised residuals of vaccine uptake (**a**, **b**) and other placebo outcomes (**c**-**f**) for different regional regression models. Additional analyses checked for explicit confounding by considering the outcome of second dose uptake as of 1 October 2021 while controlling for trust in expert sources of COVID-19 information (**a**, **b**), while placebo analyses considered various placebo outcomes that are expected to be influenced by trust in expert authority or willingness to follow expert advice but *not* by misinformation exposure—namely the proportion of obese or overweight adult population (**c**, **d**) and the proportion of physically active adult population (**e**, **f**)—both before (**a**, **c**, **e**) and after (**b**, **d**, **f**) controlling for spatial effects of regional adjacency. All models include regional real and fake news detection ability scores as predictors, while controlling for a set of regional covariates. Main plots show the posterior mean of the standardised residuals (on the *y*-axis) against the posterior mean of the outcome (on the *x*-axis)—we observe no apparent trend in the residuals suggesting homoscedasticity. Inset plots show the quantile-quantile plot for the standard normal distribution on the *x*-axis and for the (posterior mean of) standardised residuals on the *y*-axis, with the reference line in red—we observe a good fit suggesting that residuals are normally distributed.

Tables S1 to S22

Table S1. Variable names for survey and census data. For census variable names refer to the corresponding data dictionaries [\[S29–](#page-90-5)[S31\]](#page-90-6). For survey variable names see *[Questionnaire](#page-20-0)*. For variable recodes refer to Tables [S2](#page-48-0) and [S5.](#page-51-0)

Table S3. Survey sample is representative across most of the considered socio-demographic covariates. Survey counts (*n*) indicate the number of samples in each covariate group. Census percentages (%) have been computed from the 2011 census microdata [\[S29–](#page-90-5)[S31\]](#page-90-6). The "Great Britain" column excludes survey and census samples from Northern Ireland.

Table S4. Survey sample is representative across most of the geographies at the first administrative level. Population (Pop) percentages have been computed from the 2019 population estimates [\[S9,](#page-88-0) [S28\]](#page-90-7). See Fig. [S10](#page-33-0) for a map of NUTS-1 geographies.

NUTS-1 code	NUTS-1 name	Survey		Pop
		n	$\%$	$\%$
UKC	North East (England)	1512	9.18	3.99
UKD	North West (England)	2440	14.81	10.95
UKE	Yorkshire and The Humber	1337	8.11	8.23
UKF	East Midlands (England)	1465	8.89	7.22
UKG	West Midlands (England)	1508	9.15	8.86
UKH	East of England	1639	9.95	9.36
UKI	London	1289	7.82	13.48
UKJ	South East (England)	2217	13.46	13.77
UKK	South West (England)	1828	11.09	8.43
UKL	Wales	489	2.97	4.71
UKM	Scotland	445	2.70	8.18
UKN	Northern Ireland	308	1.87	2.83

Table S5. Recoding of region variables in census data to grouped NUTS-3 regions in the United Kingdom (UK). Survey respondents were mapped to grouped NUTS-3 regions via their outward postcode [\[S38,](#page-90-8) [S39\]](#page-90-9). For grouped NUTS-3 region names, and a breakdown of survey respondents across regions, refer to Table [S6.](#page-58-0)

Northern Ireland (LA_CODE_2014) [\[S31\]](#page-90-6)

Table S6. Number of survey samples across 149 grouped NUTS-3 regions of the United Kingdom (UK). Population (Pop) percentages have been computed from the 2019 population estimates [\[S9,](#page-88-0) [S28\]](#page-90-7). For brevity, results elsewhere use only the first of the NUTS-3 codes for those regions consisting of multiple NUTS-3 regions. See Fig. [S10](#page-33-0) for a map of NUTS-3 regions.

Covariate	Group	Real	Fake
	18-24	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	25-34	0.21(0.06, 0.36)	0.37(0.19, 0.57)
	35-44	0.46(0.23, 0.7)	1.06(0.6, 1.56)
Age	45-54	0.88(0.5, 1.32)	1.78(1.04, 2.63)
	55-64	0.97(0.55, 1.45)	2.49(1.43, 3.65)
	$65+$	0.72(0.38, 1.09)	3.0(1.74, 4.4)
	Level-0	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Level-1	0.51(0.24, 0.78)	0.55(0.28, 0.86)
	Level-2	0.9(0.47, 1.32)	0.94(0.53, 1.41)
Education	Level-3	1.2(0.64, 1.75)	1.62(0.92, 2.39)
	Level-4	1.72(0.93, 2.49)	1.73(1.0, 2.56)
	Other	0.9(0.48, 1.33)	0.91(0.5, 1.37)
	Null	0.29 (-0.02 , 0.64)	0.29(0.03, 0.6)
Employment	Employed	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Unemployed	0.03 ($-0.1, 0.17$)	-0.03 $(-0.16, 0.09)$
	Student	0.59(0.29, 0.9)	0.37(0.17, 0.59)
	Retired	0.36(0.16, 0.55)	0.62(0.31, 0.95)
	Other	0.04 (-0.07 , 0.15)	0.24(0.1, 0.39)
	Null	0.45(0.07, 0.88)	0.49(0.13, 0.89)
Gender	Female	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Male	0.23(0.12, 0.33)	-0.0 $(-0.03, 0.03)$
	White	0.0(0.0, 0.0)	0.0(0.0, 0.0)
Ethnicity	Other	-0.25 $(-0.4, -0.1)$	-0.57 $(-0.84, -0.31)$
	Null	-0.38 $(-0.77, -0.03)$	$-0.17(-0.49, 0.12)$
	Christian	0.0(0.0, 0.0)	0.0(0.0, 0.0)
Religion	Atheist	0.86(0.5, 1.27)	0.81(0.46, 1.2)
	Other	-0.13 $(-0.25, -0.03)$	-0.14 $(-0.25, -0.03)$
	Null	-0.44 $(-0.67, -0.21)$	-0.14 $(-0.28, -0.02)$
Region	Population (Density)	$-0.11 (-0.25, 0.01)$	-0.13 $(-0.26, -0.01)$
	Female (Proportion)	$0.07 (-0.02, 0.17)$	$0.07 (-0.01, 0.16)$
	Age 60+ (Proportion)	-0.05 $(-0.21, 0.1)$	0.03 (-0.12 , 0.17)
	Income (Per head)	-0.01 $(-0.17, 0.13)$	$0.07 (-0.08, 0.21)$
	Life expectancy (60-64)	-0.09 $(-0.23, 0.05)$	$-0.06(-0.21, 0.07)$
	Higher degree (Proportion)	0.18(0.04, 0.33)	0.04 (-0.08 , 0.17)

Table S7. Posterior estimates of covariate parameters for the social-IRT model for the United Kingdom (UK). Values indicate posterior means with 95% HPDIs in parentheses. Values in bold (underline) indicate that the HPDI is above (below) the reference group's value of 0. "Null" corresponds to undisclosed socio-demographic identity; see Table [S2](#page-48-0) for full variable recodes.

Table S8. Model convergence diagnostics for key parameters of the (primary) social-IRT models for the United Kingdom (UK). The potential scale reduction factor satisfies $\hat{R} \le 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-11).

Table S9. Posterior estimates of covariate parameters for the secondary social-IRT model for Great Britain (GB). Values indicate posterior means with 95% HPDI in parentheses. Values in bold (underline) indicate that the HPDI is above (below) the reference group's value of 0. "Null" corresponds to undisclosed socio-demographic identity; see Table [S2](#page-48-0) for full variable recodes.

Covariate	Group	Real	Fake
	18-24	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	25-34	0.21(0.06, 0.38)	0.43(0.23, 0.66)
	35-44	0.51(0.28, 0.74)	1.26(0.72, 1.8)
Age	45-54	0.99(0.6, 1.39)	2.16(1.26, 3.1)
	55-64	1.11(0.68, 1.56)	3.04(1.81, 4.35)
	$65+$	0.84(0.49, 1.21)	3.64(2.18, 5.24)
	Level-0	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Level-1	0.55(0.28, 0.81)	0.64(0.33, 0.99)
Education	Level-2	0.96(0.57, 1.36)	1.12(0.65, 1.66)
	Level-3	1.26(0.78, 1.8)	1.94(1.13, 2.81)
	Level-4	1.83(1.14, 2.56)	2.0(1.19, 2.91)
	Other	0.95(0.55, 1.35)	1.05(0.59, 1.55)
	Null	0.45(0.09, 0.83)	0.2 (-0.11 , 0.55)
	Employed	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Unemployed	0.17(0.01, 0.34)	0.1 (-0.05 , 0.27)
	Student	0.77(0.45, 1.13)	0.47(0.23, 0.74)
	Retired	0.47(0.26, 0.7)	0.87(0.47, 1.29)
Employment Gender	Other	0.14(0.01, 0.27)	0.4(0.2, 0.62)
	Null	0.7(0.23, 1.21)	0.58(0.14, 1.09)
	Female	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Male	0.23(0.14, 0.33)	0.0 (-0.04 , 0.04)
	White	0.0(0.0, 0.0)	0.0(0.0, 0.0)
Ethnicity	Asian	-0.56 $(-0.83, -0.3)$	-0.4 $(-0.65, -0.18)$
	Black	$0.04 (-0.17, 0.25)$	-0.9 $(-1.32, -0.48)$
	Mixed	-0.25 $(-0.51, -0.02)$	-0.84 $(-1.28, -0.45)$
	Other	0.29 (-0.06 , 0.63)	-0.32 $(-0.69, -0.03)$
	Null	-0.33 $(-0.74, 0.03)$	-0.26 $(-0.65, 0.12)$
	Christian	0.0(0.0, 0.0)	0.0(0.0, 0.0)
	Atheist	1.0(0.61, 1.38)	1.0(0.58, 1.43)
Religion	Muslim	0.02 (-0.18 , 0.23)	-0.57 $(-0.88, -0.28)$
	Other	-0.04 $(-0.17, 0.07)$	-0.11 $(-0.26, 0.01)$
		-0.42 $(-0.64, -0.22)$	$-0.19(-0.36, -0.03)$
	Level-0	0.0(0.0, 0.0)	0.0(0.0, 0.0)
Income	Null 0.2(0.09, 0.32) Level-1	0.27(0.13, 0.43)	
	Level-2	0.45(0.27, 0.66)	0.43(0.24, 0.66)
	Null	$-0.15(-0.3, -0.01)$	0.35(0.15, 0.56)
	Population (Density)	-0.06 $(-0.2, 0.07)$	$-0.18(-0.36, -0.03)$
	Female (Proportion)	$0.08(-0.02, 0.17)$	0.09 (-0.01 , 0.21)
	Age 60+ (Proportion)	-0.02 $(-0.2, 0.15)$	-0.02 $(-0.23, 0.18)$
Region	Income (Per head)	-0.05 $(-0.21, 0.12)$	$0.04 (-0.13, 0.22)$
	Unemployment $(\%)$	-0.07 $(-0.15, 0.02)$	$-0.01 (-0.1, 0.08)$
	Life expectancy (60-64)	-0.02 $(-0.2, 0.13)$	-0.1 $(-0.29, 0.08)$
	Voting (EU Remain)	0.21(0.04, 0.4)	-0.07 $(-0.26, 0.13)$
	Higher degree (Proportion)	$0.0 (-0.21, 0.23)$	$0.13(-0.11, 0.39)$

Table S10. Model convergence diagnostics for key parameters of the (secondary) social-IRT models for Great Britain (GB). The potential scale reduction factor satisfies $\hat{R} \leq 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-11).

M ${}^{\mu}_{r}, M$
For t $\frac{\mu}{f}, M$ $_{\rm v}^\mu,M$ $\frac{\mu}{d}, M$ n^{μ} are 5,5,10,0.88,0.88, respectively. Values in **bold** (<u>underline</u>) indicate that the HPDI is above (below) the reference value. For full NUTS-3 codes and names, see Table [S6](#page-58-1) and Fig. [S10.](#page-33-1)

M_n^μ	M_d^{μ}	M_{ν}^{μ}	M_f^{μ}	M_r^{μ}	Grouped NUTS-3 code
0.2(0.12, 0.27)	2.32(1.97, 2.68)	12.31 (11.88, 12.75)	7.22(6.93, 7.51)	5.09(4.76, 5.41)	UKC11
0.17(0.1, 0.24)	2.41(2.05, 2.77)	12.79 (12.39, 13.27)	7.52(7.24, 7.79)	5.28(4.95, 5.61)	UKC12
0.15(0.11, 0.19)	2.33(2.11, 2.55)	13.48 (13.21, 13.74)	7.83(7.68, 8.0)	5.65(5.44, 5.86)	UKC13
0.08(0.05, 0.12)	2.71(2.39, 3.06)	14.01 (13.65, 14.39)	8.32(8.13, 8.52)	5.69(5.38, 6.0)	UKC21
0.18(0.14, 0.23)	2.26(2.05, 2.48)	13.07(12.81, 13.33)	7.58(7.4, 7.74)	5.49 (5.29, 5.69)	UKC22
0.12(0.07, 0.18)	2.7(2.3, 3.11)	12.59(12.11, 13.04)	7.59(7.29, 7.87)	5.01(4.63, 5.36)	UKC23
0.08(0.05, 0.11)	2.79(2.5, 3.07)	13.64 (13.34, 13.97)	8.17(8.0, 8.35)	5.47(5.21, 5.73)	UKD11
0.4(0.28, 0.53)	1.59(1.3, 1.88)	12.7(12.29, 13.11)	6.95(6.67, 7.24)	5.76(5.46, 6.05)	UKD33
0.23(0.17, 0.29)	2.08(1.84, 2.31)	13.06 (12.77, 13.34)	7.45(7.27, 7.63)	5.61(5.39, 5.82)	UKD34
0.16(0.11, 0.21)	2.34(2.06, 2.61)	13.27 (12.95, 13.58)	7.72(7.51, 7.91)	5.54(5.29, 5.79)	UKD35
0.14(0.1, 0.19)	2.57(2.32, 2.81)	12.38 (12.11, 12.67)	7.4(7.22, 7.58)	4.97(4.75, 5.19)	UKD36
0.22(0.17, 0.27)	2.19(1.96, 2.38)	12.5(12.24, 12.76)	7.23(7.06, 7.41)	5.27(5.07, 5.47)	UKD37
0.1(0.04, 0.17)	2.84(2.32, 3.37)	12.38 (11.78, 12.96)	7.56(7.18, 7.93)	4.82 (4.37, 5.29)	UKD41
0.25(0.12, 0.4)	2.07(1.56, 2.59)	12.65 (11.97, 13.28)	7.24(6.8, 7.67)	5.42 (4.93, 5.92)	UKD42
0.14(0.1, 0.18)	2.42(2.17, 2.65)	13.58 (13.31, 13.86)	7.93(7.76, 8.1)	5.65(5.43, 5.87)	UKD44
0.23(0.15, 0.31)	2.09(1.78, 2.38)	12.98 (12.6, 13.36)	7.42(7.18, 7.67)	5.56 (5.26, 5.84)	UKD46
0.21(0.13, 0.3)	2.08(1.71, 2.44)	13.46 (13.02, 13.92)	7.67(7.39, 7.96)	5.79 (5.44, 6.14)	UKD47
0.28(0.18, 0.4)	1.8(1.45, 2.13)	13.63(13.17, 14.08)	7.57(7.28, 7.85)	6.06(5.7, 6.39)	UKD61
0.11(0.07, 0.15)	2.54(2.2, 2.85)	13.83 (13.46, 14.2)	8.13(7.91, 8.33)	5.7(5.4, 6.01)	UKD62
0.19(0.12, 0.26)	2.12(1.82, 2.45)	13.81 (13.43, 14.21)	7.87(7.65, 8.11)	5.93 (5.63, 6.24)	UKD63
0.22(0.15, 0.29)	2.14(1.86, 2.44)	12.81 (12.44, 13.17)	7.36(7.13, 7.6)	5.44(5.17, 5.72)	UKD71
0.38(0.27, 0.48)	1.7(1.42, 1.95)	12.5(12.12, 12.86)	6.91(6.66, 7.17)	5.59(5.31, 5.85)	UKD72
0.14(0.07, 0.2)	2.46(2.07, 2.85)	13.42 (12.97, 13.87)	7.87(7.6, 8.14)	5.54(5.16, 5.89)	UKD73
0.24(0.16, 0.34)	1.92(1.6, 2.24)	13.71 (13.28, 14.11)	7.69(7.45, 7.95)	6.02(5.69, 6.34)	UKD74
0.14(0.07, 0.21)	2.6(2.14, 3.0)	12.47 (11.96, 12.97)	7.47(7.12, 7.78)	5.0(4.62, 5.39)	UKE11
0.1(0.06, 0.14)	2.7(2.38, 3.03)	13.57 (13.2, 13.92)	8.09 (7.88, 8.29)	5.48 (5.19, 5.78)	UKE12

 $M_r^{\mu}, M_f^{\mu}, M_v^{\mu}, M_d^{\mu}$ are re-centered around the posterior means of UKD44 (5.65), UKE44 (7.56), UKH31 (13.21), UKI61 (2.14), and UKL13 (0.22), respectively. Values in **bold** (<u>underline</u>) indicate that the HPDI is above (below) the re-centered reference value of 0. For full NUTS-3 codes and names, see Table [S6](#page-58-1) and Fig. [S10.](#page-33-1)

Grouped NUTS-3 code	M_r^{μ}	M_f^{μ}	M_{ν}^{μ}	M_d^{μ}	M_n^{μ}
UKC11	-0.56 $(-0.96, -0.18)$	-0.34 $(-0.67, -0.0)$	-0.9 $(-1.56, -0.24)$	$0.18(-0.33, 0.64)$	-0.03 $(-0.16, 0.09)$
UKC12	$-0.37 (-0.75, 0.02)$	-0.04 $(-0.36, 0.29)$	-0.42 $(-1.08, 0.22)$	$0.27 (-0.24, 0.75)$	-0.06 $(-0.18, 0.06)$
UKC13	$0.0 (-0.27, 0.31)$	0.27(0.04, 0.49)	$0.27(-0.31, 0.87)$	0.2 (-0.23 , 0.61)	$-0.07 (-0.18, 0.03)$
UKC21	$0.04 (-0.32, 0.4)$	0.77(0.51, 1.03)	0.8(0.16, 1.45)	0.58(0.07, 1.05)	$-0.14 (-0.24, -0.04)$
UKC22	$-0.16(-0.42, 0.12)$	0.02 (-0.2 , 0.25)	$-0.14 (-0.68, 0.41)$	$0.13 (-0.29, 0.53)$	-0.04 $(-0.14, 0.06)$
UKC23	-0.64 $(-1.09, -0.23)$	0.03 (-0.28 , 0.37)	-0.62 $(-1.24, 0.04)$	0.56(0.05, 1.07)	$-0.1 (-0.22, 0.0)$
UKD11	$-0.18(-0.49, 0.13)$	0.62(0.38, 0.86)	0.43 (-0.2 , 1.03)	0.65(0.18, 1.1)	-0.14 $(-0.24, -0.05)$
UKD33	$0.11 (-0.24, 0.45)$	-0.61 $(-0.94, -0.3)$	$-0.51(-1.2, 0.16)$	-0.55 $(-1.02, -0.07)$	0.18(0.02, 0.33)
UKD34	-0.04 $(-0.36, 0.24)$	-0.1 $(-0.35, 0.14)$	-0.15 $(-0.72, 0.42)$	-0.06 $(-0.47, 0.33)$	$0.0(-0.11, 0.11)$
UKD35	$-0.11(-0.43, 0.21)$	$0.17(-0.1, 0.41)$	0.06 (-0.48 , 0.62)	$0.2 (-0.21, 0.62)$	-0.06 $(-0.17, 0.04)$
UKD36	-0.68 $(-0.97, -0.38)$	-0.15 $(-0.39, 0.08)$	-0.83 $(-1.39, -0.29)$	0.44 (-0.01 , 0.87)	-0.08 $(-0.19, 0.02)$
UKD37	-0.38 $(-0.67, -0.08)$	-0.32 $(-0.55, -0.09)$	$-0.71(-1.26, -0.14)$	$0.05 (-0.34, 0.43)$	$-0.01 (-0.12, 0.1)$
UKD41	-0.83 $(-1.36, -0.34)$	$0.01 (-0.41, 0.4)$	-0.83 $(-1.61, -0.0)$	0.71(0.09, 1.36)	$-0.12(-0.24, 0.0)$
UKD42	$-0.23(-0.75, 0.31)$	-0.32 $(-0.75, 0.18)$	$-0.56(-1.29, 0.23)$	$-0.06(-0.7, 0.58)$	0.03 (-0.14 , 0.2)
UKD44	0.0(0.0, 0.0)	0.37(0.15, 0.6)	$0.37(-0.2, 0.97)$	$0.28(-0.14, 0.72)$	$-0.09(-0.19, 0.01)$
UKD46	$-0.09(-0.45, 0.27)$	-0.13 $(-0.44, 0.14)$	$-0.23 (-0.87, 0.41)$	$-0.04 (-0.51, 0.43)$	0.0 (-0.12, 0.12)
UKD47	$0.14 (-0.25, 0.51)$	$0.11 (-0.23, 0.42)$	0.25 (-0.46 , 0.91)	-0.05 $(-0.55, 0.46)$	$-0.01 (-0.14, 0.11)$
UKD61	0.41 ($-0.0, 0.79$)	0.02 (-0.32 , 0.35)	$0.42 (-0.22, 1.08)$	$-0.33(-0.81, 0.16)$	0.06 (-0.09 , 0.2)
UKD62	$0.05(-0.31, 0.4)$	0.57(0.32, 0.84)	0.62(0.02, 1.25)	0.4 (-0.04 , 0.87)	-0.12 $(-0.22, -0.02)$
UKD63	$0.28(-0.06, 0.65)$	0.32(0.03, 0.6)	$0.6(-0.04, 1.23)$	-0.01 $(-0.46, 0.41)$	-0.04 $(-0.16, 0.07)$
UKD71	$-0.21 (-0.57, 0.15)$	$-0.19(-0.47, 0.1)$	-0.4 $(-0.97, 0.19)$	$0.01 (-0.42, 0.43)$	-0.0 $(-0.13, 0.12)$
UKD72	-0.06 $(-0.4, 0.27)$	-0.65 $(-0.94, -0.35)$	$-0.71(-1.29, -0.09)$	-0.44 $(-0.88, 0.02)$	0.15(0.01, 0.29)
UKD73	$-0.11 (-0.51, 0.3)$	0.32(0.01, 0.62)	$0.21 (-0.41, 0.86)$	$0.33(-0.14, 0.8)$	-0.09 $(-0.2, 0.03)$
UKD74	0.36(0.01, 0.75)	$0.14 (-0.16, 0.43)$	0.5 (-0.11 , 1.08)	-0.22 $(-0.66, 0.22)$	$0.02(-0.1, 0.15)$
UKE11	-0.65 $(-1.09, -0.17)$	-0.09 $(-0.46, 0.28)$	-0.74 $(-1.42, -0.04)$	$0.47(-0.1, 1.05)$	$-0.09(-0.21, 0.04)$

Table S13. MIST scores exhibit multicollinearity. When predicting a COVID-19 vaccine's second dose uptake rates y in England and Scotland using MIST scores while controlling for regional covariates—of the form $y = X\beta + \varepsilon$ where X is the full design matrix obtained by concatenating the regional MIST scores with regional covariates and ε encodes the errors—values show the ordinary least squares estimate $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ for MIST scores included in the model. The $R^2(\mathbf{y}, \hat{\mathbf{y}})$ value corresponds to the proportion of variance explained by the model, with predicted values $\hat{y} = X\hat{\beta}$: see Eq. [S35.](#page-13-0) The condition number $\kappa(X)$ refers to the square root of the ratio of largest and smallest eigenvalues of **X^TX**, with larger values indicating **X^TX** is "closer" to being non-invertible i.e. to having a smallest eigenvalue of 0, which is suggestive of instability in the estimate $\hat{\beta}$. Values computed to beyond $\pm 10^8$ are indicated by $\pm \infty$. Evidently, models including (a) $\{M_r^{\mu}, M_f^{\mu}\}$ $f(x)$, M_v^{μ} } as predictors, or (b) 4 or more from $\{M_r^{\mu}, M_f^{\mu}\}$ $^{\mu}_{f}, M^{\mu}_{v}, M^{\mu}_{d}$ d_i^{μ} , M_n^{μ} } as predictors, are perfectly ill-conditioned ($\kappa(\mathbf{X}) = \infty$), since these are sets of linearly dependent variables (see Eqs.13c and [S22\)](#page-9-0). Amongst the models using 2 MIST scores as predictors, $\{M_r^{\mu}, M_f^{\mu}\}$ f_f^{μ} is the least collinear with the smallest $\kappa(X) = 8.83$. See Fig. [S12](#page-35-0) for pairwise relationships between MIST scores.

Table S14. Dominance analysis [\[S18\]](#page-89-0) suggests dominance of the fake news detection score M_f^{μ} $^{\boldsymbol \mu}_f$ over other MIST scores, when predicting regional vaccine uptake rates while controlling for regional covariates. Values indicate the difference, up to 3 significant digits, in $R²$ values of predictions between the subset model (row) and the model which includes an additional predictor (column), with values in bold indicating the largest increment in variance explained when adding a predictor to a given subset model (see Table [S13\)](#page-76-0). All models control regional covariates. Adding M_f^{μ} f ^{μ} as a predictor explains the most additional variance than adding any other MIST score to the base model with only controls. A consequence of linear dependence of $\{M_r^{\mu}, M_f^{\mu}\}$ f^{μ} , M^{μ}_{ν} } is that for every subset model consisting of M^{μ}_{f} *f* , the scores M_r^{μ} and M_v^{μ} explain identical amounts of additional variance, i.e. dominance analysis alone cannot choose which one of the two can be used alongside M_f^{μ} when predicting uptake.

Table S16. Model convergence diagnostics for key parameters when predicting COVID-19 vaccine uptake rates in England and Scotland. Non-spatial models do not control for spatial autocorrelations, while spatial models control for regional adjacency structure. The potential scale reduction factor satisfies $\hat{R} \le 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-0).

Table S17. Posterior estimates of standardized coefficients for regional MIST scores and regional covariates when predicting COVID-19 vaccine uptake rates in England and Scotland, and other placebo outcomes in England. Values indicate posterior means with 95% HPDI in parentheses. For MIST scores (real M_r^μ and fake M_f^μ f_f^{μ} news detection ability scores) and regional covariates, values in **bold** (underline) indicate that the HPDI is above (below) the ROPE of $(-0.05, 0.05)$ and the reference value of 0, respectively.

Table S18. Model convergence diagnostics of standardized coefficients for regional MIST scores and regional covariates when predicting COVID-19 vaccine uptake rates in England and Scotland, and other placebo outcomes in England. The potential scale reduction factor satisfies $\hat{R} \le 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-0).

Table S19. Posterior estimates and model convergence diagnostics for key parameters of the individual-level model of trust in an expert source of COVID-19 information for the United Kingdom (UK). Values indicate posterior means of the coefficients with 95% HPDI in parentheses, with values in **bold** (underline) indicating that the HPDI is above (below) the reference value of 0. The potential scale reduction factor satisfies $\hat{R} \le 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-0).

^Mf has ^a credible effect on self-reported vaccination status even after controlling for mere willingness to trust exper^t advice*T*.

		\widehat{R}			S_{eff}			
		M, T , controls	M , controls		Only controls M, T , controls	M , controls	Only controls	
MIST scores M	M_r				8001	19955		
	M_f		$\mathbf{1}$	$\overline{}$	7455	18009		
	$M_r \times T$				8001	19955		
	$M_f \times T$		$\mathbf{1}$	$\overline{}$	7455	18009		
Trust T	Some expert			\overline{a}	8001			
	No expert	$\mathbf{1}$	$\overline{}$	$\overline{}$	11493			
Age	18-24				11493			
	$25 - 34$		$\mathbf{1}$	$\mathbf{1}$	7856	8232	8229	
	35-44				8439	8364	8278	
	45-54		$\mathbf{1}$	1	8227	8615	8410	
	55-64				7939	8330	8226	
	$65+$		$\mathbf{1}$	$\mathbf{1}$	7804	7910	7756	
Education	Level-0		1		7804	7910	7756	
	Level-1		$\mathbf{1}$	1	13460	13269	8328	
	Level-2				14011	14370	8679	
	Level-3		$\mathbf{1}$	$\mathbf{1}$	15420	15767	8265	
	Level-4				14013	12463	4678	
	Other		$\mathbf{1}$	$\mathbf{1}$	13809	15930	9264	
	Null		$\mathbf{1}$	$\mathbf{1}$	13556	15781	11775	
Employment	Employed		$\mathbf{1}$		13556	15781	11775	
	Unemployed		$\mathbf{1}$	$\mathbf{1}$	8825	8104	8116	
	Student		1	1	11352	11431	11059	
	Retired		$\mathbf{1}$	$\mathbf{1}$	7055	7515	7519	
	Other				8121	7809	8050	
	Null		$\mathbf{1}$	$\mathbf{1}$	7623	7352	8061	

 $\hat{R} \le 1.02$ and the effective sample size satisfies $S_{\text{eff}} > 400$ [\[S40\]](#page-90-1).

Table S22. Momentum resampling diagnostics for the social-IRT, regional level and individual level models. The Bayesian fraction of missing information [\[S41\]](#page-90-2) satisfies BFMI > 0.3 [\[S42\]](#page-90-3) across all 4 chains of all models.

Model name	Model type	Chain 1	Chain 2	Chain 3	Chain 4						
Social-IRT models											
United Kingdom (primary)	Real	0.83	0.81	0.80	0.78						
	Fake	0.72	0.83	0.74	0.81						
Great Britain (secondary)	Real	0.85	0.79	0.79	0.77						
	Fake	0.75	0.72	0.77	0.76						
Regional level models											
Second dose uptake (primary)	Non-spatial	0.94	0.92	0.95	0.94						
as of 1 October 2021	Spatial	0.44	0.42	0.47	0.46						
First dose uptake (secondary)	Non-spatial	0.88	0.94	0.96	0.91						
as of 1 October 2021	Spatial	0.39	0.34	0.40	0.42						
First dose uptake (secondary)	Non-spatial	0.94	0.94	0.98	0.97						
as of 1 July 2021	Spatial	0.42	0.45	0.42	0.39						
Second dose uptake (robustness)	Non-spatial	0.87	0.94	0.90	0.95						
controlling for trust in experts	Spatial	0.40	0.44	0.45	0.42						
Overweight or obese (placebo)	Non-spatial	0.98	0.92	0.91	0.99						
	Spatial	0.63	0.65	0.60	0.70						
Physical activity (placebo)	Non-spatial	1.00	0.90	1.01	0.93						
	Spatial	0.68	0.60	0.63	0.60						
Individual-level models											
Vaccine uptake (robustness)	M, T , controls	0.83	0.76	0.82	0.78						
	M , controls	0.76	0.81	0.79	0.77						
	Only controls	0.77	0.74	0.78	0.81						
Trust in expert info sources		0.78	0.75	0.8	0.79						

Supplementary references

- S1. Loomba, S. *et al. Regional relationship of misinformation susceptibility & vaccine uptake: Preregistration of Planned Analyses* <https://osf.io/um2hd/>. Nov. 2021.
- S2. NHS England. *COVID-19 vaccination programme: FAQs on second doses* [https://](https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2021/03/C1254-covid-19-vaccination-programme-faqs-on-second-dose-v2.pdf) [www.england.nhs.uk/coronavirus/wp- content/uploads/sites/](https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2021/03/C1254-covid-19-vaccination-programme-faqs-on-second-dose-v2.pdf) [52/2021/03/C1254-covid-19-vaccination-programme-faqs-on](https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2021/03/C1254-covid-19-vaccination-programme-faqs-on-second-dose-v2.pdf)[second-dose-v2.pdf](https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2021/03/C1254-covid-19-vaccination-programme-faqs-on-second-dose-v2.pdf). Accessed: 2022-05-04. 2021.
- S3. Riebler, A., Sørbye, S. H., Simpson, D. & Rue, H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 25, 1145–1165. doi:[10.1177/0962280216660421](https://doi.org/10.1177/0962280216660421) (2016).
- S4. Facebook Data for Good. *Facebook Social Connectedness Index* [https://data.](https://data.humdata.org/dataset/social-connectedness-index) [humdata.org/dataset/social-connectedness-index](https://data.humdata.org/dataset/social-connectedness-index). Accessed: 2021- 10-23. 2021.
- S5. Besag, J., York, J. & Mollie, A. Bayesian image restoration, with two applications in ´ spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, 1–20. doi:[10.](https://doi.org/10.1007/BF00116466) [1007/BF00116466](https://doi.org/10.1007/BF00116466) (1991).
- S6. Snijders, T. A. Statistical Models for Social Networks. *Annual Review of Sociology* 37, 131–153. doi:[10.1146/annurev.soc.012809.102709](https://doi.org/10.1146/annurev.soc.012809.102709) (2011).
- S7. Hoff, P. D., Raftery, A. E. & Handcock, M. S. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association* 97, 1090–1098. doi:[10.](https://doi.org/10.1198/016214502388618906) [1198/016214502388618906](https://doi.org/10.1198/016214502388618906) (2002).
- S8. Loomba, S. & Jones, N. S. *Geodesic statistics for random network families* 2021. doi:[10.48550/arXiv.2111.02330](https://doi.org/10.48550/arXiv.2111.02330).
- S9. Eurostat. *Population by age group, sex and NUTS 3 region* [https://ec.europa.](https://ec.europa.eu/eurostat/web/products-datasets/product?code=demo_r_pjangrp3) eu / eurostat / web / products - datasets / product ? code = demo_r_ [pjangrp3](https://ec.europa.eu/eurostat/web/products-datasets/product?code=demo_r_pjangrp3). Accessed: 2021-10-23. 2021.
- S10. Office for National Statistics. *Annual Survey of Hours and Earnings Resident Analysis* [https://www.nomisweb.co.uk/query/construct/summary.asp?](https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=30) [mode=construct&version=0&dataset=30](https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=30). Accessed: 2021-11-11. 2021.
- S11. Burr, I. W. Cumulative Frequency Functions. *The Annals of Mathematical Statistics* 13, 215–232. doi:[10.1214/aoms/1177731607](https://doi.org/10.1214/aoms/1177731607) (1942).
- S12. Rodriguez, R. N. A guide to the Burr type XII distributions. *Biometrika* 64, 129–134. ISSN: 0006-3444. doi:[10.1093/biomet/64.1.129](https://doi.org/10.1093/biomet/64.1.129) (Apr. 1977).
- S13. Office for National Statistics. *Annual Population Survey* [https://www.nomisweb.](https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=17) [co.uk/query/construct/summary.asp?mode=construct&version=](https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=17) [0&dataset=17](https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=17). Accessed: 2021-11-11. 2021.
- S14. Gelman, A. & Little, T. C. Poststratification Into Many Categories Using Hierarchical Logistic Regression. *Survey Methodology* 23, 127–135. doi:[10.1.1.44.5270](https://doi.org/10.1.1.44.5270) (1997).
- S15. Leemann, L. & Wasserfallen, F. Extending the Use and Prediction Precision of Subnational Public Opinion Estimation. *American Journal of Political Science* 61, 1003–1022. doi:[10.1111/ajps.12319](https://doi.org/10.1111/ajps.12319) (2017).
- S16. Hoover, J. & Dehghani, M. The big, the bad, and the ugly: Geographic estimation with flawed psychological data. *Psychological Methods* 25, 412. doi:[10 . 1037 /](https://doi.org/10.1037/met0000240) [met0000240](https://doi.org/10.1037/met0000240) (2020).
- S17. Jensen, J. L. W. V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30, 175–193. doi:[10.1007/BF02418571](https://doi.org/10.1007/BF02418571) (1906).
- S18. Budescu, D. V. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin* 114, 542. doi:[10.](https://doi.org/10.1037/0033-2909.114.3.542) [1037/0033-2909.114.3.542](https://doi.org/10.1037/0033-2909.114.3.542) (1993).
- S19. Azen, R. & Budescu, D. V. The Dominance Analysis Approach for Comparing Predictors in Multiple Regression. *Psychological Methods* 8, 129. doi:[10 . 1037 / 1082 -](https://doi.org/10.1037/1082-989X.8.2.129) [989X.8.2.129](https://doi.org/10.1037/1082-989X.8.2.129) (2003).
- S20. Moran, P. A. P. Notes on Continuous Stochastic Phenomena. *Biometrika* 37, 17–23. ISSN: 0006-3444. doi:[10.1093/biomet/37.1-2.17](https://doi.org/10.1093/biomet/37.1-2.17) (June 1950).
- S21. Rey, S. J. & Anselin, L. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies* 37, 5–27 (2007).
- S22. Maertens, R. *et al.* The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods.* doi:[10.](https://doi.org/10.3758/s13428-023-02124-2) [3758/s13428-023-02124-2](https://doi.org/10.3758/s13428-023-02124-2) (2023 in press).
- S23. Roozenbeek, J. *et al.* Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making* 17. doi:[10 . 1017 /](https://doi.org/10.1017/S1930297500003570) [S1930297500003570](https://doi.org/10.1017/S1930297500003570) (May 2022).
- S24. Kruschke, J. K. Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science* 1, 270–280. doi:[10.1177/](https://doi.org/10.1177/2515245918771304) [2515245918771304](https://doi.org/10.1177/2515245918771304) (2018).
- S25. Eggers, A. C., Tuñón, G. & Dafoe, A. *Placebo Tests for Causal Inference* [https :](https://pelg.ucsd.edu/Eggers_2021.pdf) [//pelg.ucsd.edu/Eggers_2021.pdf](https://pelg.ucsd.edu/Eggers_2021.pdf). Accessed: 2022-10-25. 2021.
- S26. Jones, D. E., Carson, K. A., Bleich, S. N. & Cooper, L. A. Patient trust in physicians and adoption of lifestyle behaviors to control high blood pressure. *Patient Education and Counseling* 89, 57–62. ISSN: 0738-3991. doi:[10.1016/j.pec.2012.06.003](https://doi.org/10.1016/j.pec.2012.06.003) (2012).
- S27. Eurostat. *United Kingdom NUTS level 3 (2016)* [https : / / ec . europa . eu /](https://ec.europa.eu/eurostat/documents/345175/7451602/2016-NUTS-3-map-UK.pdf) [eurostat/documents/345175/7451602/2016-NUTS-3-map-UK.pdf](https://ec.europa.eu/eurostat/documents/345175/7451602/2016-NUTS-3-map-UK.pdf). Accessed: 2021-10-01. 2018.
- S28. Office for National Statistics. *Local Authority District (December 2018) to NUTS3 to NUTS2 to NUTS1 (January 2018) Lookup in United Kingdom* [https : / /](https://geoportal.statistics.gov.uk/datasets/ons::local-authority-district-december-2018-to-nuts3-to-nuts2-to-nuts1-january-2018-lookup-in-united-kingdom/about) [geoportal.statistics.gov.uk/datasets/ons::local-authority](https://geoportal.statistics.gov.uk/datasets/ons::local-authority-district-december-2018-to-nuts3-to-nuts2-to-nuts1-january-2018-lookup-in-united-kingdom/about) $district - december - 2018 - to - nuts3 - to - nuts2 - to - nuts1$ [january-2018-lookup-in-united-kingdom/about](https://geoportal.statistics.gov.uk/datasets/ons::local-authority-district-december-2018-to-nuts3-to-nuts2-to-nuts1-january-2018-lookup-in-united-kingdom/about). 2019.
- S29. Office for National Statistics. *2011 Census Microdata Individual Safeguarded Sample (Local Authority): England and Wales* 2015. doi:[10.5255/UKDA-SN-7682-1](https://doi.org/10.5255/UKDA-SN-7682-1).
- S30. National Records of Scotland. *2011 Census Microdata Individual Safeguarded Sample (Local Authority): Scotland* 2015. doi:[10.5255/UKDA-SN-7835-1](https://doi.org/10.5255/UKDA-SN-7835-1).
- S31. Northern Ireland Statistics and Research Agency. *2011 Census Microdata Individual Safeguarded Sample (Local Authority): Northern Ireland* 2015. doi:[10.5255/UKDA-](https://doi.org/10.5255/UKDA-SN-7770-1)[SN-7770-1](https://doi.org/10.5255/UKDA-SN-7770-1).
- S32. Oliphant, T. E. *A Bayesian perspective on estimating mean, variance, and standarddeviation from data* 2006. <http://hdl.lib.byu.edu/1877/438>.
- S33. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17, 261–272. doi:[10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2) (2020).
- S34. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21, 1–13. doi:[10.1186/s12864-019-6413-7](https://doi.org/10.1186/s12864-019-6413-7) (2020).
- S35. Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* 32, 241–254. doi:[10.](https://doi.org/10.1007/BF02289588) [1007/BF02289588](https://doi.org/10.1007/BF02289588) (1967).
- S36. Lance, G. N. & Williams, W. T. A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Journal* 9, 373–380. ISSN: 0010-4620. doi:[10.](https://doi.org/10.1093/comjnl/9.4.373) [1093/comjnl/9.4.373](https://doi.org/10.1093/comjnl/9.4.373) (Feb. 1967).
- S37. Public Health England. *Coronavirus in the UK* [https://coronavirus.data.](https://coronavirus.data.gov.uk/details/about-data) [gov.uk/details/about-data](https://coronavirus.data.gov.uk/details/about-data). Accessed: 2021-10-23. 2021.
- S38. Eurostat. *Postcodes and NUTS* https://ec.europa.eu/eurostat/web/ [nuts/correspondence-tables/postcodes-and-nuts](https://ec.europa.eu/eurostat/web/nuts/correspondence-tables/postcodes-and-nuts). Accessed: 2021- 10-01. 2021.
- S39. Eurostat. *Changes from NUTS 2016 to NUTS 2021* [https : / / ec . europa . eu /](https://ec.europa.eu/eurostat/documents/345175/7451602/NUTS3-changes-2016-to-2021.pdf) [eurostat/documents/345175/7451602/NUTS3- changes- 2016- to-](https://ec.europa.eu/eurostat/documents/345175/7451602/NUTS3-changes-2016-to-2021.pdf)[2021.pdf](https://ec.europa.eu/eurostat/documents/345175/7451602/NUTS3-changes-2016-to-2021.pdf). Accessed: 2021-10-01. 2021.
- S40. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved *R* for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis* 16, 667–718. doi:[10 . 1214 / 20 -](https://doi.org/10.1214/20-BA1221) [BA1221](https://doi.org/10.1214/20-BA1221) (2021).
- S41. Betancourt, M. *Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo* 2016. doi:[10.48550/arXiv.1604.00695](https://doi.org/10.48550/arXiv.1604.00695).
- S42. Betancourt, M. *A Conceptual Introduction to Hamiltonian Monte Carlo* 2018. doi:[10.](https://doi.org/10.48550/arXiv.1701.02434) [48550/arXiv.1701.02434](https://doi.org/10.48550/arXiv.1701.02434).