Supplementary Materials

The effects of communicating scientific uncertainty on trust and decision making in a public health context

Claudia R. Schneider[1,2], Alexandra L. J. Freeman[1], David Spiegelhalter[1], Sander van der Linden[1,2]

[1]Winton Centre for Risk and Evidence Communication, University of Cambridge

[2]Department of Psychology, University of Cambridge

Correspondence to: Claudia R. Schneider, Department of Psychology and Winton Centre for Risk and Evidence Communication, University of Cambridge, Cambridge, CB2 3EB, United Kingdom. E-mail: cs2025@cam.ac.uk

Table of Contents

Study 1

Scenario/Experimental wording

Control group ('Control'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 8 will die. This is known as the COVID-19 case fatality rate.*

<u>Low quality of evidence conditions:</u>

Low QoE group – Disagreement ('Low-disagree'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 8 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is uncertain, because there is disagreement between experts.

Low QoE group - Lack of data ('Low-lack'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 8 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is uncertain, because there is a lack of data.

Screenshot of the Oxford COVID-19 Evidence Service website

We used data from the Oxford COVID-19 Evidence Service by the Oxford Centre for Evidence-Based Medicine to provide participants with a current estimate of the case fatality rate for the UK at the time of each study. As the page has been continuously updated, the historic information is not available anymore; we hence include screenshots of the page at the time of Study 1 below (screenshots of the page as of April 2nd 2020; retrieved using the wayback machine internet archive, https://web.archive.org/web/20200402114744/https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/). In addition to the case fatality rate, the page also provides information on the uncertainty about the case fatality rate.

CEBM

The Centre for Evidence-Based Medicine develops, promotes and disseminates better evidence for healthcare.

HOME COVID-19 EVIDENCE SCHOOL OF EBHC BLOG RESOURCES OUTREACH

Global Covid-19 Case Fatality Rates

March 17, 2020

CEBM RESEARCH Oxford COVID-19 Evidence Service

Jason Oke, Carl Heneghan
Updated 31st March

Lay Summary by Mandy Payne, Health Watch

This page is updated daily as new information emerges. It sets out the current Case Fatality Rate (CFR) estimates, the country-specific issues affecting the CFR, and provides a current best estimate of the CFR, and more importantly, the Infection Fatality Rate (IFR).

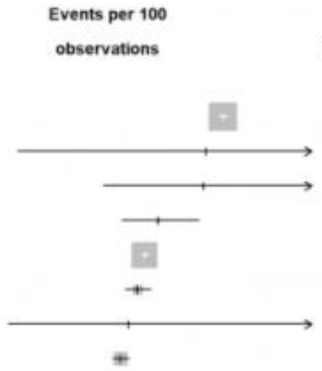
The IFR estimates the fatality rate in all those with infection: the detected disease (cases) and those with an undetected disease (asymptomatic and not tested group).

Case Fatality Rates:

The total number of cases and the total number of deaths from COVID-19 outbreak data was drawn down (scraped) from https://www.worldometers.info/coronavirus/.

The proportion of deaths to the total numbers of cases was meta-analysed using the R function *metaprop*, using fixed-effect inverse-variance weighting. Estimates from the cruise ship 'Diamond Princess' as well as countries with three or fewer deaths to date recorded are excluded from the analysis. We present country-level case fatality as a percentage along with 95% confidence intervals in a forest plot. Estimates of heterogeneity and a 95% prediction interval are presented, but a pooled overall estimate is suppressed due to heterogeneity. (understanding data in meta-analysis)

*case fatality rate is the number of reported deaths per number of reported cases (Updated 31st March)



Country	Deaths	Cases	Case Fatality (%)	95%-CI
Italy	12428	105792	11.75	(11.55 to 11.94)
Bangladesh	6	54	11.11	(4.19 to 22.63)
San Marino	26	236	11.02	(7.32 to 15.72)
Indonesia	157	1677	9.36	(8.01 to 10.86)
Spain	9053	102136	8.86	(8.69 to 9.04)
Netherlands	1173	13614	8.62	(8.15 to 9.10)
DRC	9	109	8.26	(3.85 to 15.10)
UK	2352	29474	7.98	(7.67 to 8.30)

Last updated: April 01, 2020, 15:50 GMT

Between countries, case Fatality rates vary significantly, and over time, which suggests considerable uncertainty over the exact case fatality rates. see: Prediction intervals for CFR overtime pdf

- The number of cases detected by testing will vary considerably by country;
- Selection bias can mean those with severe disease are preferentially tested;
- There may be delays between symptoms onset and deaths which can lead to underestimation of the CFR;
- There may be factors that account for increased death rates such as coinfection, more inadequate healthcare, patient demographics (i.e., older patients might be more prevalent in countries such as Italy);
- There may be increased rates of smoking or comorbidities amongst the fatalities.
- Differences in how deaths are attributed to Coronavirus: dying with the disease (association) is not the same as dying from the disease (causation).

Measures

Mediator measure

To what extent do you think that the COVID-19 case fatality rate mentioned is certain? [1- Not certain at all, 7-Very certain]

Dependent variables

Perceived trustworthiness index items:

To what extent do you think that the COVID-19 case fatality rate mentioned is accurate? [1-Not accurate at all, 7-Very accurate]

To what extent do you think that the COVID-19 case fatality rate mentioned is reliable? [1-Not reliable at all, 7-Very reliable]

To what extent do you think the COVID-19 case fatality rate mentioned is trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

Decision making index items:

How likely are you to base your own COVID-19 related decisions and behaviours on the mentioned case fatality rate? [1-Not at all likely, 7-Very likely]

To what extent do you think the government should base its decisions on how to handle the pandemic on the mentioned COVID-19 case fatality rate? [1-Not at all, 7-Very much]

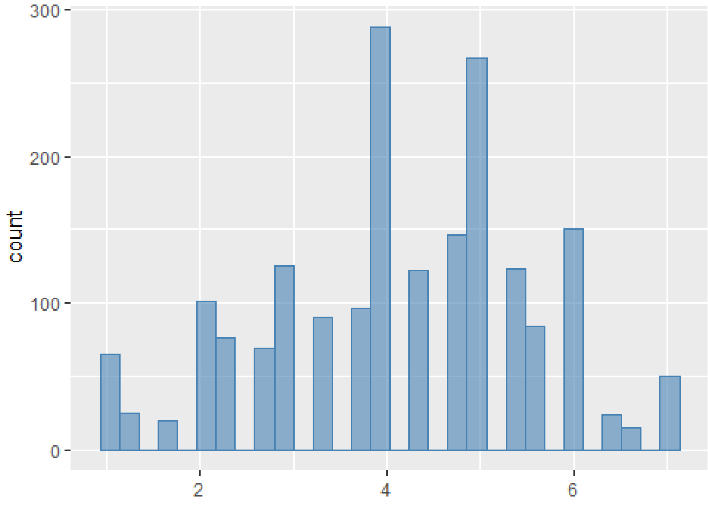
Attention check measure

What is the estimated COVID-19 case fatality rate for the UK that we showed you?

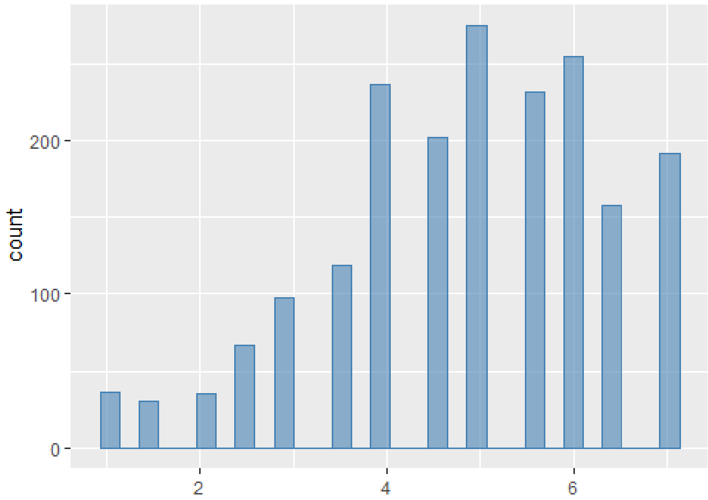
- 6 out of 100 people
- 8 out of 100 people
- 12 out of 100 people
- 4 out of 100 people

Distributions of main measures

Perceived trustworthiness:



Decision making:



Demographic composition of study sample

Table S1. Demographic characteristics of the study sample (Study 1).

Variable	**Study 1** (*N* = 1,942)
Gender, %	
Females	52.01
Males	47.68
Age, *M* (*SD*)	45.61 (15.72)
Education, %	83.01
Political views, *M* (*SD*)	3.82 (1.40)

Note: Due to some participants choosing not to indicate gender, percentages do not total to 100.

*Educational Attainment = *at least Bachelors Degree or equivalent.*

*Political views on spectrum from left wing (or liberal) to right wing (or conservative) on 7-point scale.

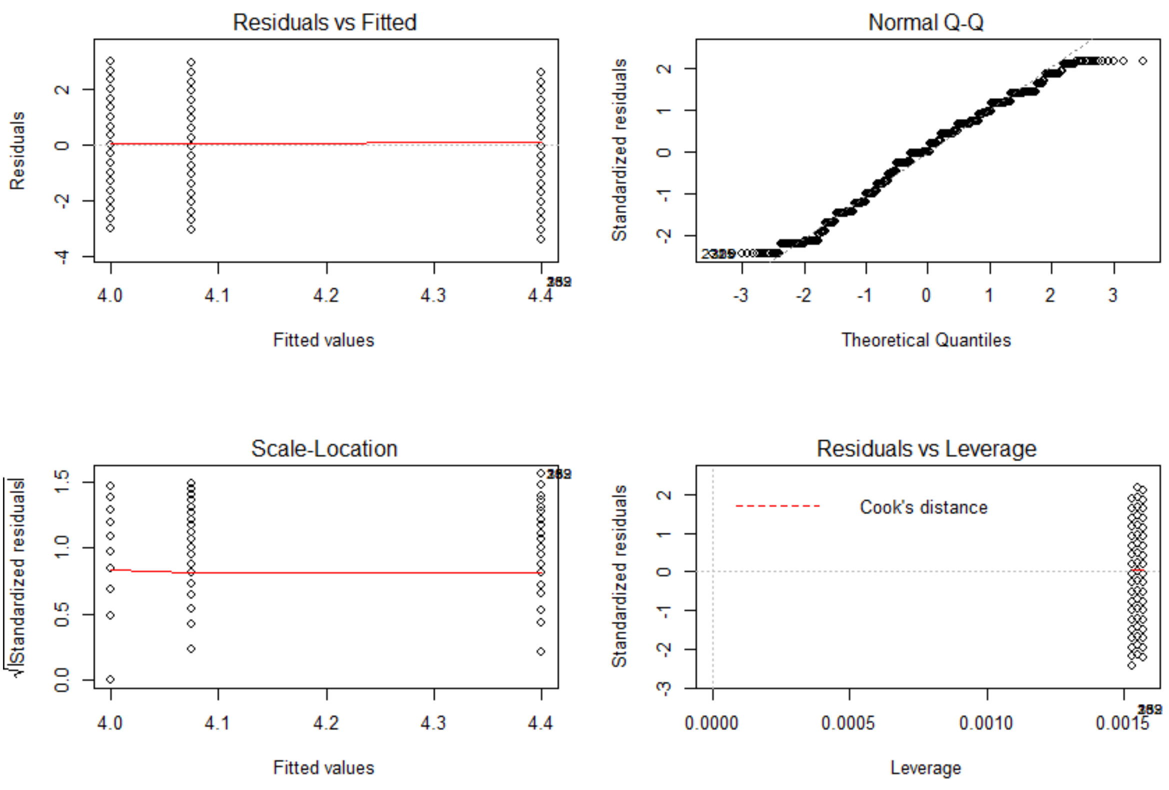
Descriptive statistics for main outcome measures

Table S2. Means and standard deviations per experimental group for perceived trustworthiness and use in decision making.

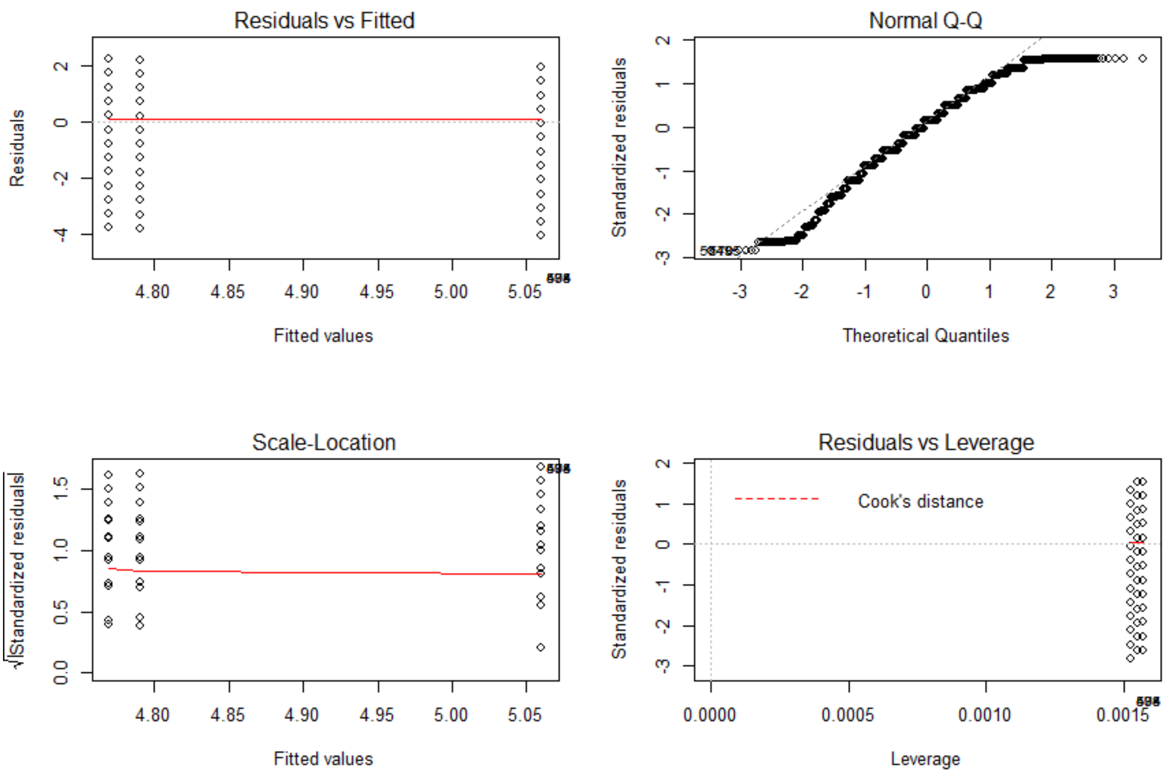
	Perceived trustworthiness		Use in decision making	
Condition	***Mean***	***SD***	***Mean***	***SD***
Control	4.40	1.38	5.06	1.39
Low-disagree	4.08	1.38	4.77	1.49
Low-lack	4.00	1.43	4.79	1.44

Model Diagnostics

Perceived trustworthiness:



Decision Making:



Experimental balance checks

We run balance checks to test whether random assignment successfully balanced demographic background variables across experimental groups, as outlined in the pre-registration. Results are detailed below.

Gender: $\chi^2(2,1936) = 0.008$, $p = .996$

Age: $F(2, 1924) = 0.44$, $p = .646$

Education (treated as categorical): $\chi^2(10,1940) = 8.89$, $p = .543$

Education (treated as continuous): $F(2, 1937) = 3.28$, $p = .038$

Politics: $F(2, 1935) = 2.22$, $p = .109$

Where random assignment did not successfully balance the variables across our experimental groups, we control for these imbalances in our models, as per our pre-registration.

Perceived trustworthiness:

Test of main effect of experimental condition controlling for education: $F(2, 1931) = 16.06$, $p < .001$, $\eta_p^2 = 0.016$

Decision making:

Test of main effect of experimental condition controlling for education: $F(2, 1931) = 8.44$, $p < .001$, $\eta_p^2 = 0.009$

Experimental effects remain significant after controlling for education.

Sampling platform check

We ran two-way analysis of variance to test for potential effects of sampling platform on our experimental results. Experimental effects stayed significant controlling for sampling platform. Additionally, we do not find any significant interactions for any of the measures.

Perceived trustworthiness:

Test of main effect of experimental condition controlling for sampling platform: $F(2, 1932) = 14.99$, $p < .001$, $\eta_p^2 = 0.015$

Test of interaction between sampling platform and experimental condition: $F(2, 1930) = 0.70$, $p = .495$

Decision making:

Test of main effect of experimental condition controlling for sampling platform: $F(2, 1932) = 8.24$, $p < .001$, $\eta_p^2 = 0.008$

Test of interaction between sampling platform and experimental condition: $F(2, 1930) = 2.83$, $p = .060$

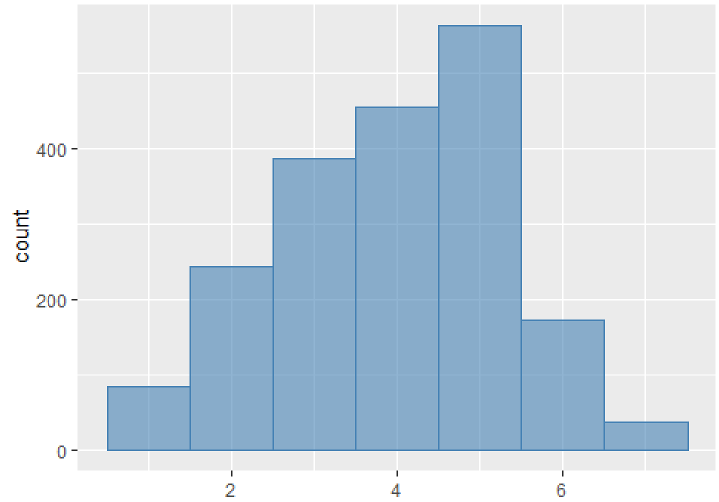
Knowledge measure analysis

Across all three studies, participants were asked how much they thought was currently known about COVID-19. This measure was found to be uninformative due to problems with its design, and hence the reporting was removed from the main paper, as suggested in the review process. Analyses are reported below. A discussion of the measure's design issues can be found in the knowledge measure section of Study 3 in the supplement.

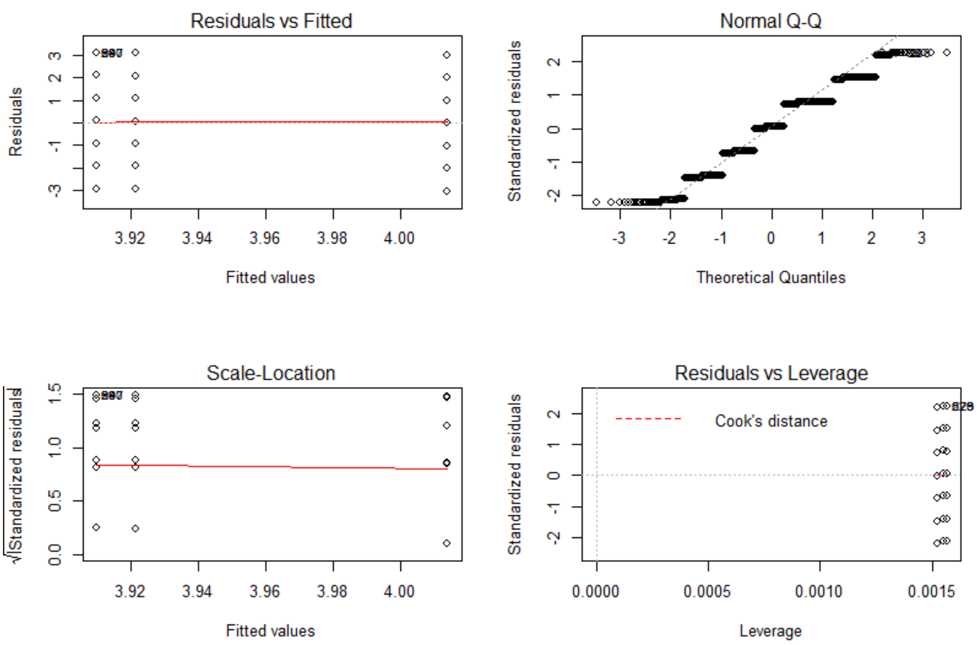
Knowledge item:

How much do you think is currently known about COVID-19? [1-Very little, 7-Very much]

Distribution of knowledge measure:



Model diagnostics:



No significant difference emerged between the experimental groups for perceived knowledge ($M_{control}$ = 4.01, $SD_{control}$ = 1.40, $M_{disagree}$ = 3.92, $SD_{disagree}$ = 1.37, M_{lack} = 3.91, SD_{lack} = 1.36; $F(2, 1937) = 1.11$, p = .331). Non-parametric robustness testing results were in line with the parametric findings (Kruskal-Wallis χ^2= 1.85, df = 2, p = .396).

The quality of evidence information provided does not seem to have an effect on perceived knowledge.

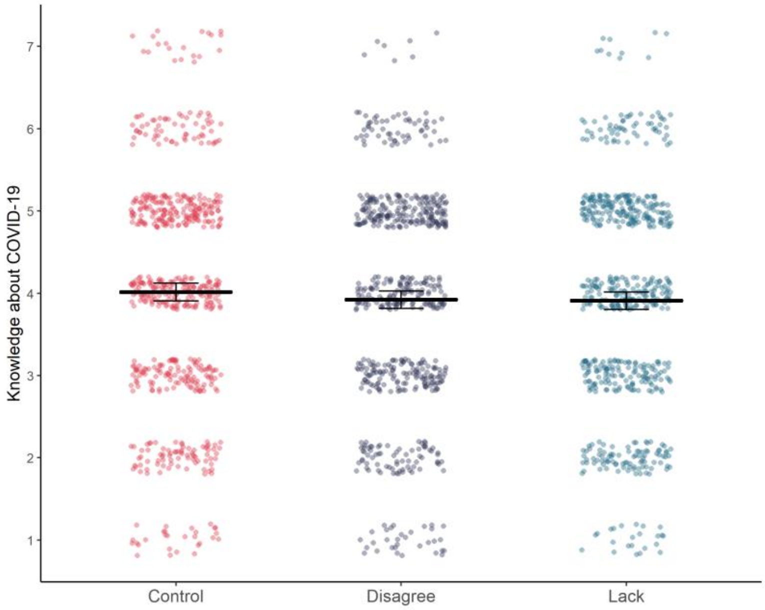


Fig S1: Experimental effects of low quality of evidence (as indicated by disagreement between experts and lack of data) versus control on knowledge about COVID-19. Error bars denote 95% confidence interval.

Direct effects analysis for SEM modelling contrast selection

Direct effects models* of experimental condition on decision making were run for all contrasts in Study 1, using the *lavaan* package in R (1,000 samples; bootstrapped confidence intervals). Significant contrasts form the basis of the structural equation modelling analysis reported in the main manuscript.

Contrast	**Estimate**	**95% CI**
Control – Low-disagree	-0.29	[-0.45, -0.14]
Control – Low-lack	-0.27	[-0.42, -0.12]
Low-disagree – Low-lack	0.02	[-0.14, 0.17]

*While we modelled latent variables for our SEM path analysis reported in the main manuscript, we were not able to include latent variables in our test of the direct effects for contrast selection, as the models in *lavaan* failed to converge for several of the contrasts. We hence base our contrast selection on direct effects modelling which does not include latent variables. Note that a comparison between those direct effects latent variable models that did converge and the respective direct effects models without latent variables revealed differences only in the estimates; however, all contrasts that were significant in the latent variables models were also significant in the models without latent variables.

Exploration of alternative path models

As outlined in the main text, we explore two alternative path models in addition to the serial mediation model presented in the paper in which manipulation of quality of evidence information affects people's perception of the certainty or uncertainty of the information, which affects perceived trustworthiness, and in turn decision making.

Alternative model 1 is also a serial mediation model, however, the sequence of the two mediators, perceived uncertainty and trustworthiness, is switched. Alternative model 2 is a parallel mediation model in which both mediators jointly affect decision making.

In the following we present and discuss findings from the alternative path models. We then compare model fit between the main serial mediation model presented in the paper and the two alternative path models respectively, using conventional rules of thumb comparisons, described in further detail below.

Note that methodological details on the employed path analysis (i.e., R package used, maximum likelihood estimation, 1000 bootstrapped samples for the confidence intervals of the indirect effects) are the same as for the serial model presented in the main text. We run the alternative models for the same contrasts as presented in the main text. Results are presented below.

Alternative model 1 (serial mediation) results:

Contrast	Indirect effect		Model fit							
	Estimate	95% CI	χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Low-disagree	-0.14	[-0.19, -0.08]	134.79	13	< .001	0.979	0.085	0.058	22740.89	22813.14
Control vs. Low-lack	-0.16	[-0.23, -0.10]	165.82	13	< .001	0.974	0.095	0.059	23109.03	23181.40

Results show that the indirect effects are significant for both investigated contrasts, however, while model fit is acceptable, fit is less good compared to the main model reported in the paper. Comparing CFI, RMSEA, and SRMR model fit indices of the alternative model 1 with fit indices of the main model presented in the paper, the alternative model 1 shows worse fit: While CFI values are above 0.95 for both contrasts, they are lower than for the main model. The same holds for SRMR values; they are below 0.08, however, they are larger compared to the main model. RMSEA values are above 0.06 which point toward potential fit issues (while RMSEA values for the main model indicate good fit).

We more formally evaluate the plausibility of the competing SEM models on the basis of Akaike (AIC) and Bayesian (BIC) Information Criteria.

According to Burnham and Anderson (2004) AIC allows for a "strength-of-evidence comparison and ranking of candidate hypotheses or models" (p. 271). In order to assess relative benefits of a set of models, i.e., quantifying meaningful differences in AIC values, the application of rules of thumb is deemed useful (Burnham and Anderson, 2004).

According to these rules of thumb a difference of ≤ 2 between models is interpreted as 'substantial support' (or evidence) that the compared models are similar in fit and plausibility, a difference of between 4 and 7 is interpreted as there being 'considerably less support' (Burnham, Anderson, & Huyvaert, 2011 indicate the range of between 2 and 7 as having 'some support'), and a difference of >10 between models is interpreted as there being 'essentially no support' that the two models are equivalent, i.e., the model with the higher AIC is the less plausible fitted model (Burnham and Anderson, 2004).

We additionally also make use of the 'Bayesian counterpart' (Burnham and Anderson, 2004, p. 271) to AIC rules of thumb for model comparison, i.e., the use of BIC values for interpreting between-model differences, first proposed by Jeffreys (1961, as cited in Raftery 1995) and adapted by Raftery (1995) for use in social science research. According to these rules of thumb a BIC difference of between 0 and 2 offers ‘weak’ evidence that the two models are different in fit (and plausibility), a BIC difference between 2 and 6 offers ‘positive’ evidence, a BIC difference between 6 and 10 offers ‘strong’ evidence, and a BIC difference of >10 offers ‘very strong’ evidence (Raftery 1995, 1999).

It should be noted that AIC and BIC are different criteria, with different properties and underlying assumptions and can hence behave differently in structural equation modelling in some circumstances (Vrieze, 2012). We report both to provide a fuller picture and note differences where occurring.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 1 (serial mediation):

Contrast	Main serial mediation model in paper		Alternative model 1 (serial mediation)		Difference (Alt1-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Low-disagree	22649.30	22721.56	22740.89	22813.14	91.59	91.58
Control vs. Low-lack	23006.82	23079.20	23109.03	23181.40	102.21	102.2

Both AIC and BIC deltas (Alt1 – Main) are positive for both contrasts, suggesting model fit of the main serial model is better compared to the alternative serial model. Furthermore, all observed AIC and BIC differences are >10 providing very strong evidence for the superiority of the main model compared to the serial alternative, following the above presented rules of thumb (Burnham and Anderson, 2004; Raftery 1995, 1999).

Alternative model 2 (parallel mediation) results:

Contrast	**Indirect effect - uncertainty**		**Indirect effect - trust**		**Model fit**							
	Estimate	**95% CI**	**Estimate**	**95% CI**	χ^2	*df*	*p*	**CFI**	**RMSEA**	**SRMR**	**AIC**	**BIC**
Control vs. Low-disagree	0.02	[-0.03, 0.06]	-0.20	[-0.32, -0.12]	41.77	11	< .001	0.995	0.047	0.016	22651.86	22734.44
Control vs. Low-lack	0.04	[-0.01, 0.11]	-0.26	[-0.38, -0.15]	59.73	11	< .001	0.992	0.058	0.015	23006.95	23089.65

For both contrasts only the indirect effect through trust emerged as significant, but not the indirect effect through uncertainty, indicating that the effect of experimental condition on decision making might not be mediated through two separate paths, and that another model (e.g., serial) might be superior. Model fit indices (CFI, RMSEA, and SRMR) are comparable to those of the main model presented in the paper.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 2 (parallel mediation):

Contrast	Main serial mediation model in paper		Alternative model 2 (parallel mediation)		Difference (Alt2-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Low-disagree	22649.30	22721.56	22651.86	22734.44	2.56	12.88
Control vs. Low-lack	23006.82	23079.20	23006.95	23089.65	0.13	10.45

Both AIC and BIC deltas (Alt2 – Main) are positive for both contrasts, suggesting model fit of the main serial model is better compared to the alternative parallel model. We observe some disagreement between AIC and BIC rules of thumb comparison of between-model differences: All observed BIC differences are >10 providing very strong evidence for the higher plausibility of the main model over the parallel model (Raftery 1995, 1999). AIC difference for the Control vs. Low-disagree contrast showed 'some support' that the main model and the parallel model are similar in fit (Burnham, Anderson, & Huyvaert, 2011); AIC difference for the Control vs. Low-lack contrast showed 'substantial support' for model equivalence (Burnham and Anderson, 2004). As noted above, given inherent differences in properties and assumptions of AIC and BIC criteria, differences can occur in structural equation modelling in some circumstances (Vrieze, 2012).

Taken together, the balance of evidence – taking both model fit as well as (non-)significance of indirect effect paths into account - points towards the main serial model being more plausible compared to the parallel alternative.

Thus, overall the evidence suggests that the main model presented in the paper is a better description of the mechanism at play, compared to the investigated alternatives.

References:

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology*, *65*(1), 23-35.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25, 111-163.

Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection". *Sociological methods & research*, *27*(3), 411-427.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*, *17*(2), 228.

Exploratory analysis of additional outcome measures

In addition to the outcome measures reported in the main text, Study 1 collected several other measures to explore a broad range of outcome variables. Most of the measures were collected again in Study 2 to replicate findings and further test the used measures, before making a final judgment on which measures to include in our confirmatory Study 3. We therefore used insights on our measures gained in studies 1 and 2 to identify the most sensitive and appropriate measures to include in Study 3. Those measures which showed repeated severe skew/ceiling effects in their distributions, or were thought to be prone to misinterpretation by the participants were ultimately dropped from the design and no longer collected in Study 3.

Results for all measures that were collected in Study 1 are reported in the following.

Overview of additional measures collected:

1. How does the information you just saw make you feel? Please indicate this by dragging the slider to select a number from 0 = negative/unhappy to 10 = positive/happy.

2. How easy or difficult do you find this information to understand? [1-Very difficult, 7-Very easy]

3. To what extent do you think the people who are responsible for calculating the COVID-19 case fatality rate mentioned are trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

4. How likely are you to provide the COVID-19 case fatality rate mentioned to a friend if they ask you for it? [1-Not at all likely, 7-Very likely]

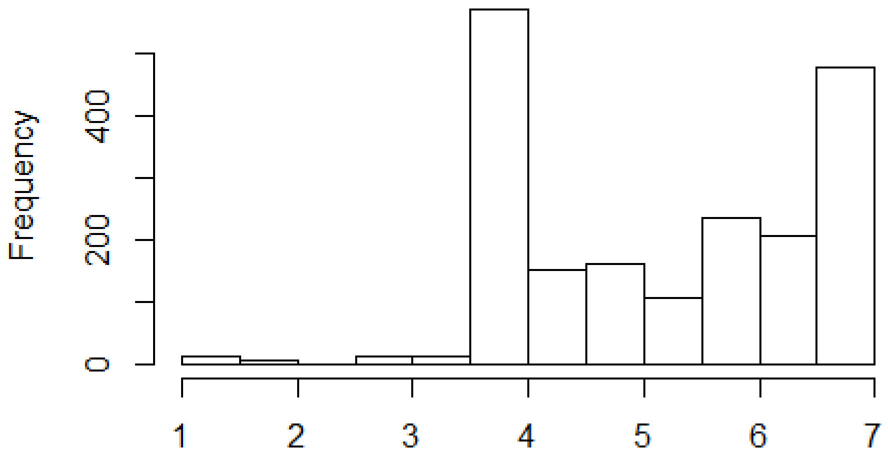
5. How strongly do you agree or disagree with the following statement: *The government should make the mentioned COVID-19 case fatality rate publicly available to everyone on its website about the disease.* [1-Strongly disagree, 7-Strongly agree]

6. Reading the COVID-19 case fatality rate information that we just showed you, how much more or less likely are you to… (1- at lot less likely, 7 – a lot more likely)

- Stay at home (following the government's lock down mandates)
- Physically distance yourself by 2 meters from others when spending time outside
- Self-isolate yourself at home for 14 days if you have recently been around someone who might be infected
- Not touch your face, including your eyes, nose, and mouth, unless your hands are clean
- Cover your nose and mouth when in public

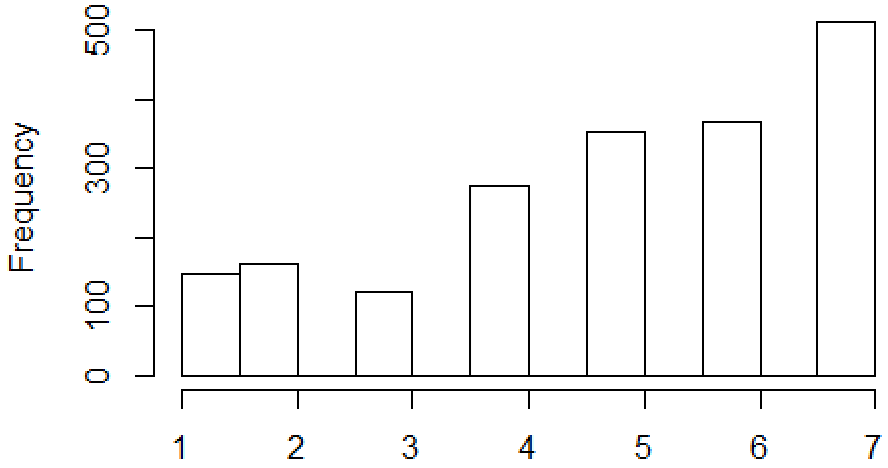
Appraisal:

Item 6: Information's effect on compliance with health mandates - The five items of the compliance with public health mandates were combined into an index (α = 0.92). Our measure of compliance with public health mandates produced a very non normal distribution with most people reporting to not have changed because of the presented information, and the remaining data showing a skew towards being a lot more likely to comply because of the shown information. We tested the measure again in Study 2 to see whether the pattern would replicate or change.

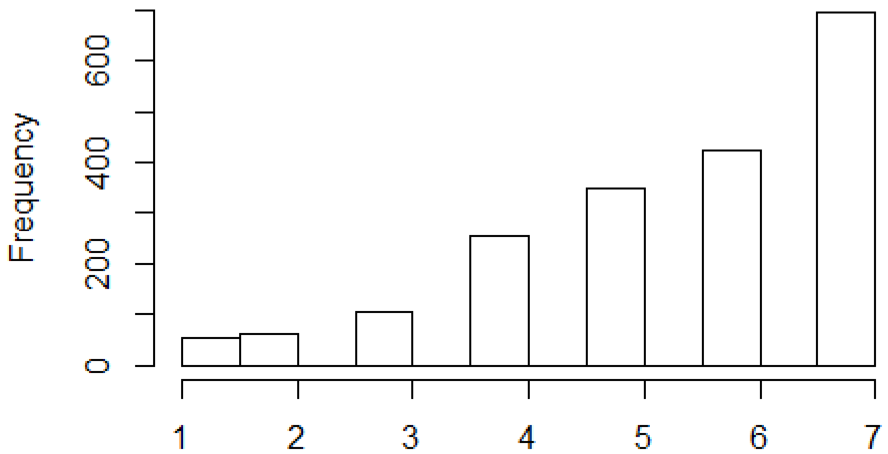


Items 4 (Sharing of information with friends) and 5 (Reporting of information by government): The distributions of these two items were somewhat skewed, especially for the government item. We tested both again in Study 2 to see whether patterns would replicate or change.

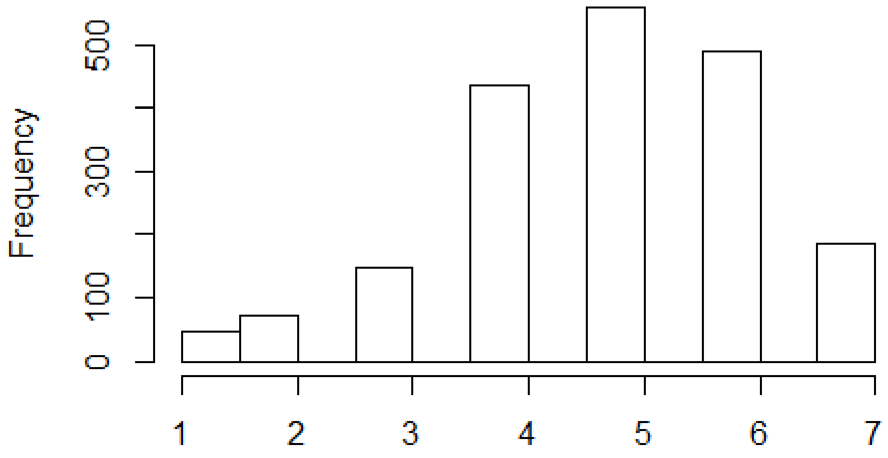
Info sharing with friends:



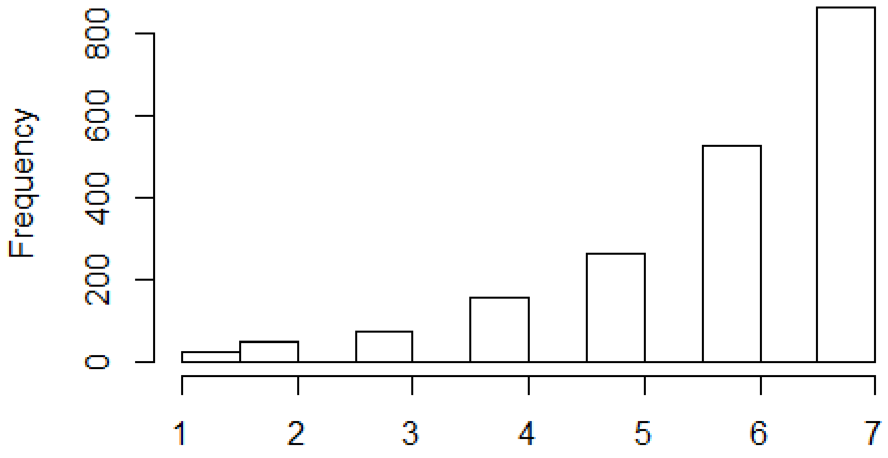
Info sharing by government:



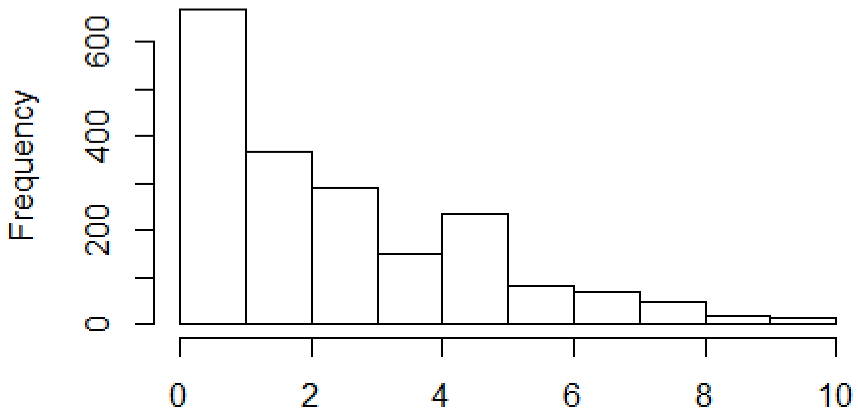
Item 3: Perceived trustworthiness of producers - This measure was kept as an exploratory outcome variable in Study 3.



Item 2: Comprehension - The measure was highly skewed in its distribution. We tested it again in Study 2 to see whether the pattern would replicate or change.



Item 1: Affect - The measure showed severe skew. We used it again in Study 2 to see whether the pattern would replicate or change.



Analysis results:

In the following we report one-way analysis of variance findings for the various outcomes. We note though that non-normality of some of the measures' distributions should be kept in mind.

Item 1: Affect

One-way analysis of variance did not find a significant effect of experimental condition on affect ($F(2, 1936) = 2.22$, $p = .109$). See visualization of descriptives (group means and confidence intervals) in Fig S2.

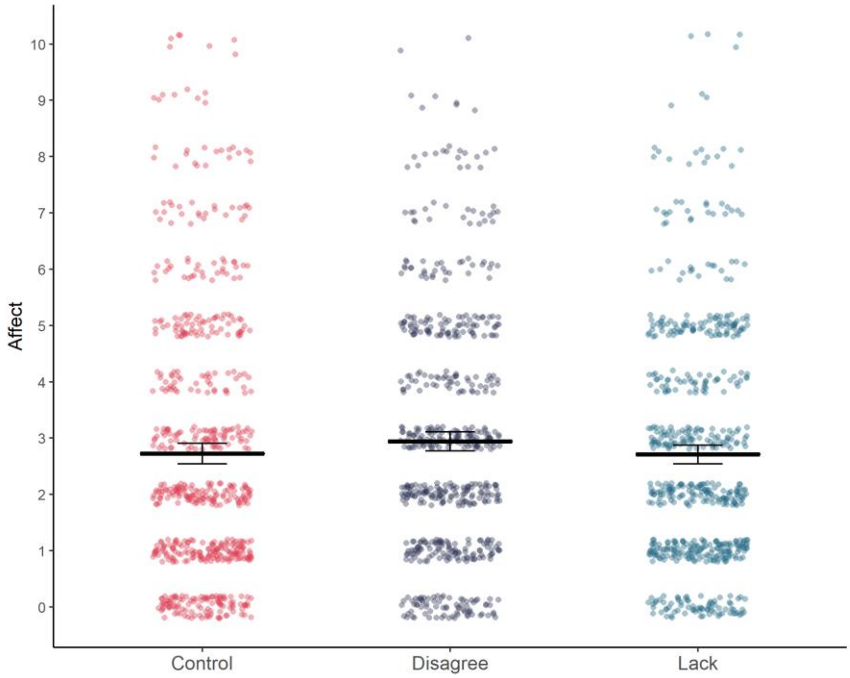


Fig. S2: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on affect. Error bars denote 95% confidence interval.

Item 2: Comprehension

One-way analysis of variance found a significant effect of experimental condition on comprehension ($F(2, 1937) = 13.14$, $p < .001$, $\eta_p^2 = 0.013$), such that the low quality of evidence groups indicated through expert disagreement as well as through lack of data reported significantly lower ease of comprehension compared to the control group respectively. Overall, comprehension for all groups was high though. See visualization of descriptives (group means and confidence intervals) in Fig S3.

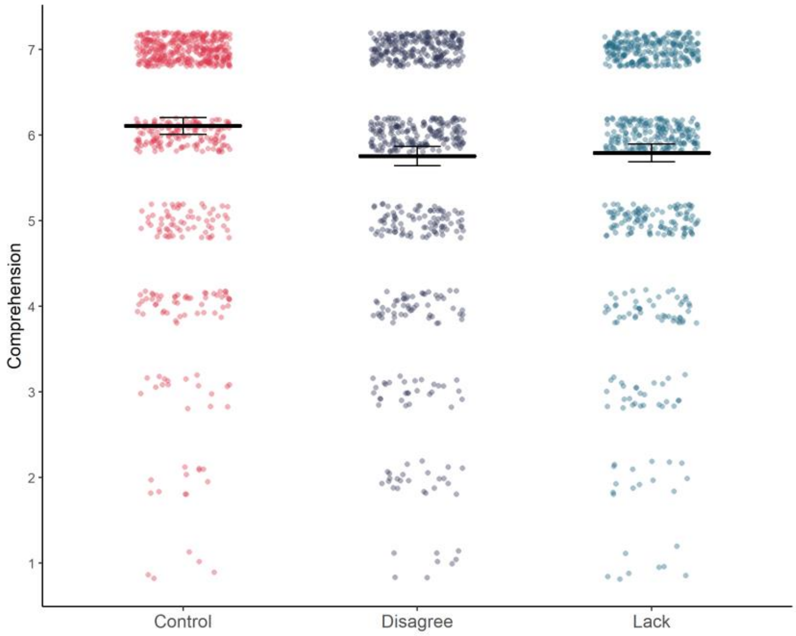


Fig. S3: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on ease of comprehension. Error bars denote 95% confidence interval.

Item 3: Perceived trustworthiness of producers

One-way analysis of variance did not find a significant effect of experimental condition on perceived trustworthiness of the producers of the information ($F(2, 1936) = 2.34$, $p = .097$). See visualization of descriptives (group means and confidence intervals) in Fig S4.

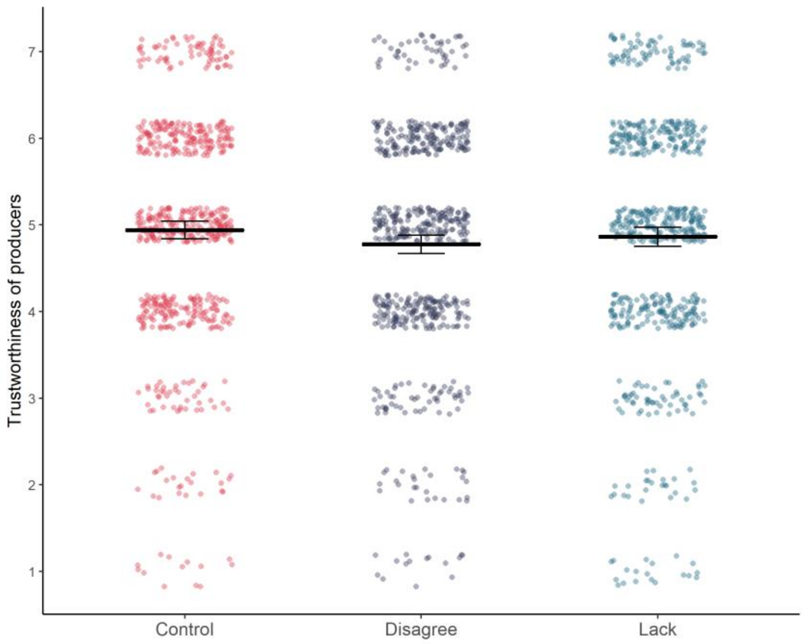


Fig. S4: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on perceived trustworthiness of the producers of the information. Error bars denote 95% confidence interval.

Item 4: Sharing of information with friends

One-way analysis of variance did not find a significant effect of experimental condition on people's inclination to share the case fatality rate information with their friends ($F(2, 1935) = 2.74$, $p = .065$). See visualization of descriptives (group means and confidence intervals) in Fig S5.

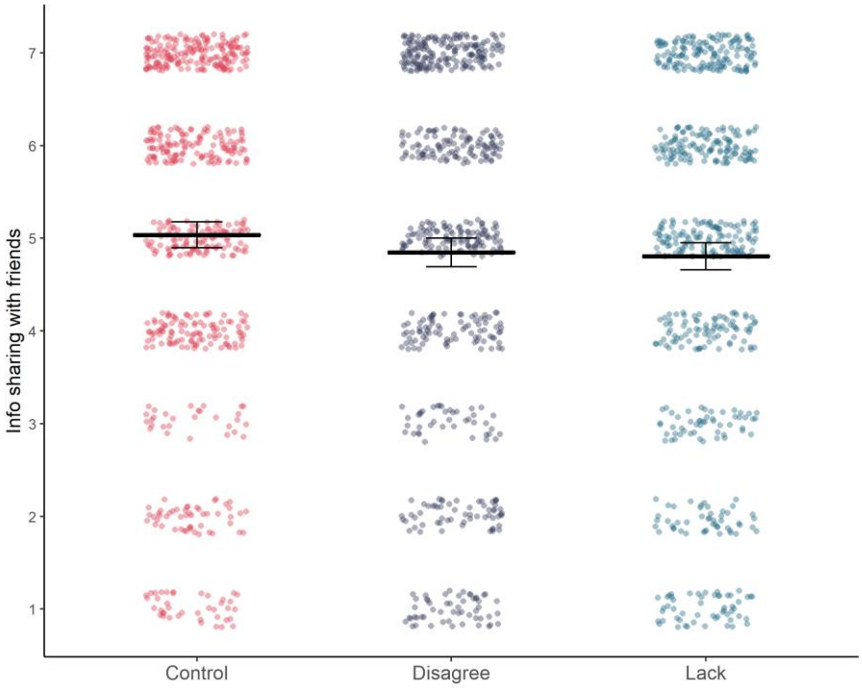


Fig. S5: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on people's intentions to share the case fatality rate information with their friends. Error bars denote 95% confidence interval.

Item 5: Reporting of information by government

One-way analysis of variance found a significant effect of experimental condition on people's support of the government making the case fatality rate information publicly available ($F(2, 1936) = 5.09$, $p = .006$, $\eta_p^2 = 0.005$), such that participants in the low quality condition as indicated by lack of data reported significantly less support compared to control group participants. See visualization of descriptives (group means and confidence intervals) in Fig S6.

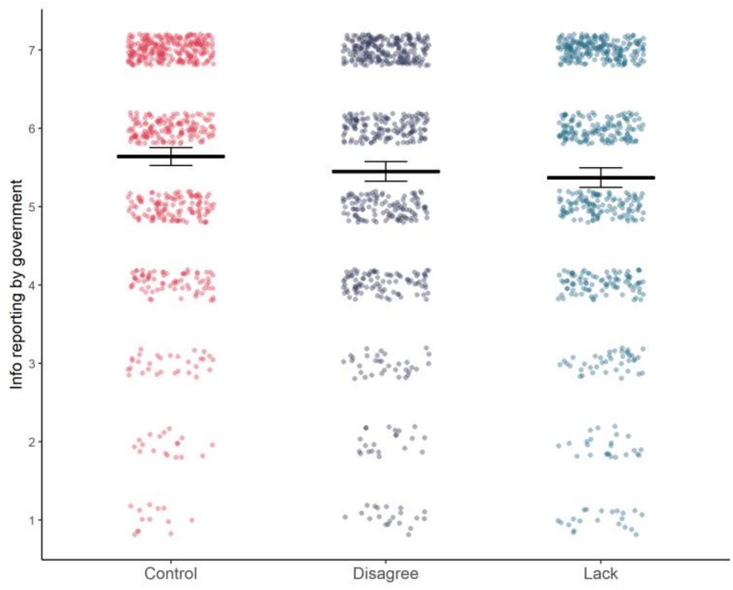


Fig. S6: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on people's support of the government making the case fatality rate information publicly available. Error bars denote 95% confidence interval.

Item 6: Information's effect on compliance with health mandates

One-way analysis of variance did not find a significant effect of experimental condition on people's reported changes in compliance with public health mandates due to the presented information ($F(2, 1937) = 0.95$, $p = .388$). See visualization of descriptives (group means and confidence intervals) in Fig S7.

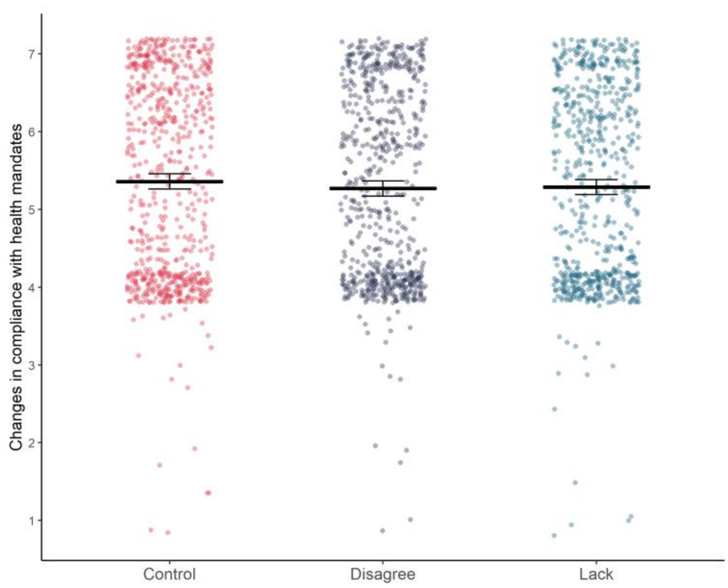


Fig. S7: Experimental effects of low quality of evidence (disagreement between experts and lack of data) versus control on people's reported changes in compliance with public health mandates. Error bars denote 95% confidence interval.

Study 2

Scenario/Experimental wording

Control group ('Control'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 15 will die. This is known as the COVID-19 case fatality rate.*

Ambiguous QoE group ('Ambiguous'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 15 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is uncertain. The quality of the evidence could be high or could be low.

<u>High quality of evidence conditions:</u>

High QoE group – Agreement ('High-agree'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 15 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is fairly certain, because there is a high level of expert agreement.

High QoE group – Data ('High-data'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 15 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is fairly certain, because there is a large amount of data available.

Screenshot of the Oxford COVID-19 Evidence Service website

We used data from the Oxford COVID-19 Evidence Service by the Oxford Centre for Evidence-Based Medicine to provide participants with a current estimate of the case fatality rate for the UK at the time of each study. As the page has been continuously updated, the historic information is not available anymore; we hence include screenshots of the page at the time of Study 2 below (screenshots of the page as of May 7th 2020; retrieved using the wayback machine internet archive, https://web.archive.org/web/20200507123543/https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/). In addition to the case fatality rate, the page also provides information on the uncertainty about the case fatality rate.

CEBM

The Centre for Evidence-Based Medicine develops, promotes and disseminates better evidence for healthcare.

HOME COVID-19 EVIDENCE SCHOOL OF EBHC BLOG RESOURCES OUTREACH

Global Covid-19 Case Fatality Rates

March 17, 2020

CEBM RESEARCH Oxford COVID-19 Evidence Service

Jason Oke, Carl Heneghan
UPDATED 7th May 2020

Lay Summary by Mandy Payne, Health Watch

This page is updated regularly as new information emerges. It sets out the current Case Fatality Rate (CFR) estimates, the country-specific issues affecting the CFR, and provides a current best estimate of the CFR, and more importantly, the Infection Fatality Rate (IFR).

The IFR estimates the fatality rate in all those with infection: the detected disease (cases) and those with an undetected disease (asymptomatic and not tested group).

Case Fatality Rates:

The total number of cases and the total number of deaths from COVID-19 outbreak data was drawn down (scraped) from https://www.worldometers.info/coronavirus/.

The proportion of deaths to the total numbers of cases was meta-analysed using the R function *metaprop*, using fixed-effect inverse-variance weighting. Estimates from the cruise ship 'Diamond Princess' as well as countries with **fewer than 1000 cases** are excluded from the analysis. (updated 9th April). We present country-level case fatality as a percentage along with 95% confidence intervals in a forest plot. Estimates of heterogeneity and a 95% prediction interval are presented, but a pooled overall estimate is suppressed due to heterogeneity. (understanding data in meta-analysis)

*case fatality rate is the number of reported deaths per number of reported cases (Updated 7th May)

Country	Deaths	Cases	Events per 100 observations	Case Fatality (%)	95%-CI
Belgium	8415	51420	>	16.37	(16.05 to 16.69)
UK	30076	201101		14.96	(14.80 to 15.11)

Last updated: May 07, 2020, 10:30 GMT

Between countries, case Fatality rates vary significantly, and over time, which suggests considerable uncertainty over the exact case fatality rates.

- The number of cases detected by testing will vary considerably by country;
- Selection bias can mean those with severe disease are preferentially tested;
- There may be delays between symptoms onset and deaths which can lead to underestimation of the CFR;
- There may be factors that account for increased death rates such as coinfection, more inadequate healthcare, patient demographics (i.e., older patients might be more prevalent in countries such as Italy);
- There may be increased rates of smoking or comorbidities amongst the fatalities.
- Differences in how deaths are attributed to Coronavirus: dying with the disease (association) is not the same as dying from the disease (causation).

Measures

Mediator measure

To what extent do you think that the COVID-19 case fatality rate mentioned is certain? [1- Not certain at all, 7-Very certain]

Dependent variables

Perceived trustworthiness index items:

To what extent do you think that the COVID-19 case fatality rate mentioned is accurate? [1-Not accurate at all, 7-Very accurate]

To what extent do you think that the COVID-19 case fatality rate mentioned is reliable? [1-Not reliable at all, 7-Very reliable]

To what extent do you think the COVID-19 case fatality rate mentioned is trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

Decision making index items:

How likely are you to base your own COVID-19 related decisions and behaviours on the mentioned case fatality rate? [1-Not at all likely, 7-Very likely]

To what extent do you think the government should base its decisions on how to handle the pandemic on the mentioned COVID-19 case fatality rate? [1-Not at all, 7-Very much]

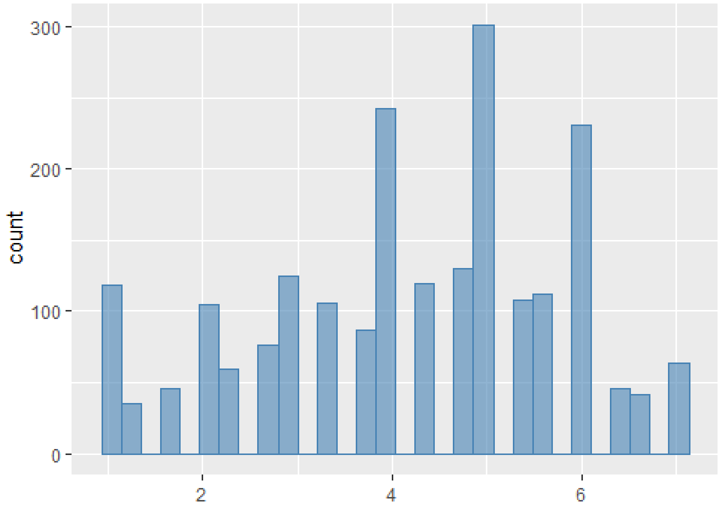
Attention check measure

What is the estimated COVID-19 case fatality rate for the UK that we showed you?

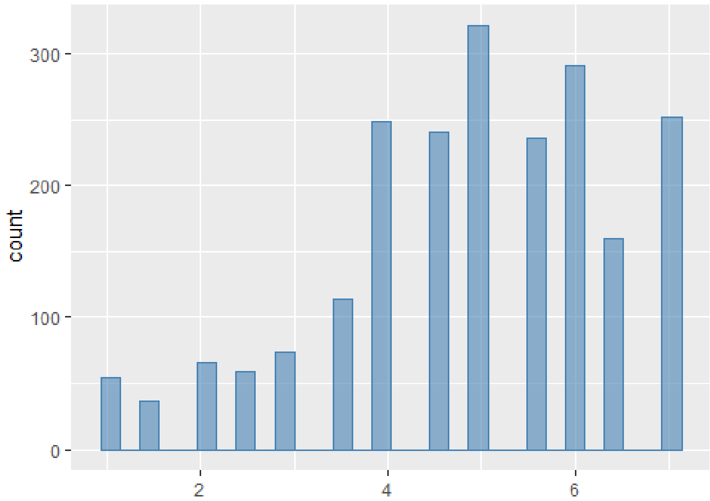
- ○ 15 out of 100 people
- ○ 12 out of 100 people
- ○ 19 out of 100 people
- ○ 8 out of 100 people

Distributions of main measures

Perceived trustworthiness:



Decision making:



Demographic composition of study sample

Table S3. Demographic characteristics of the study sample (Study 2).

Variable	**Study 2** (N = 2,155)
Gender, %	
Females	51.60
Males	48.21
Age, *M* (*SD*)	45.47 (15.76)
Education, %	84.73
Political views, *M* (*SD*)	3.78 (1.37)

Note: Due to some participants choosing not to indicate gender, percentages do not total to 100.

*Educational Attainment = *at least Bachelors Degree or equivalent.*

*Political views on spectrum from left wing (or liberal) to right wing (or conservative) on 7-point scale.

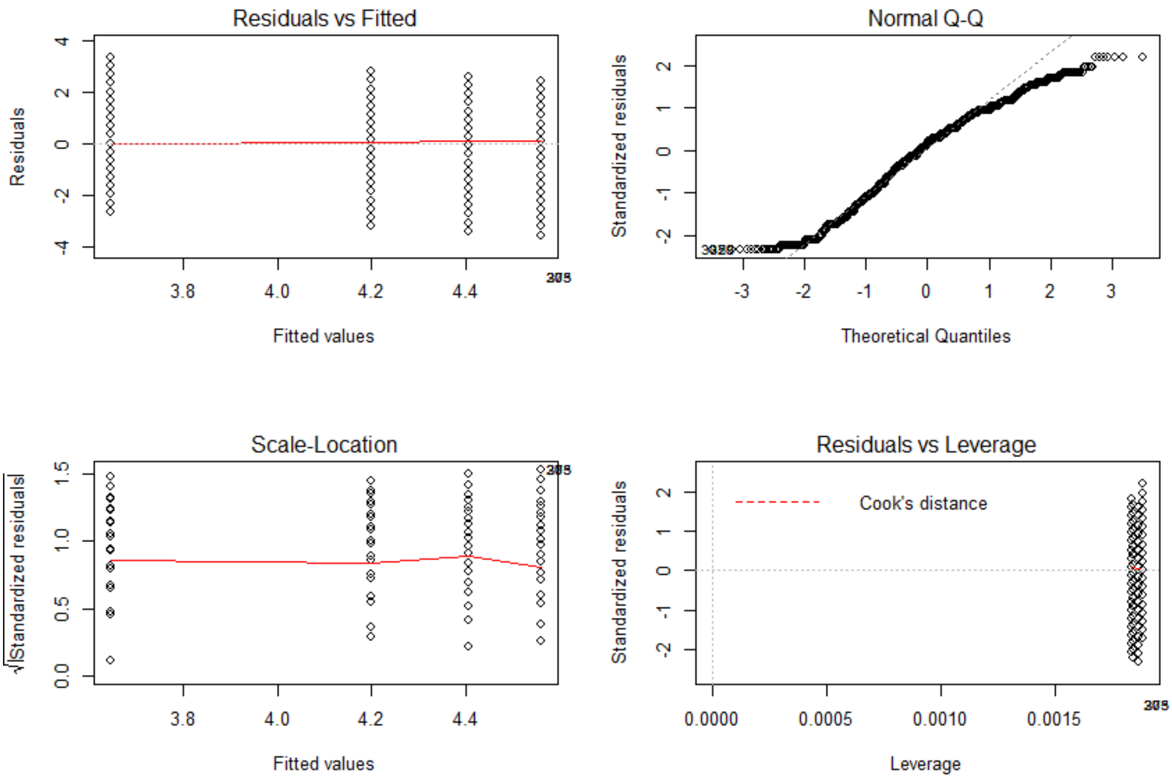
Descriptive statistics for main outcome measures

Table S4. Means and standard deviations per experimental group for perceived trustworthiness and use in decision making.

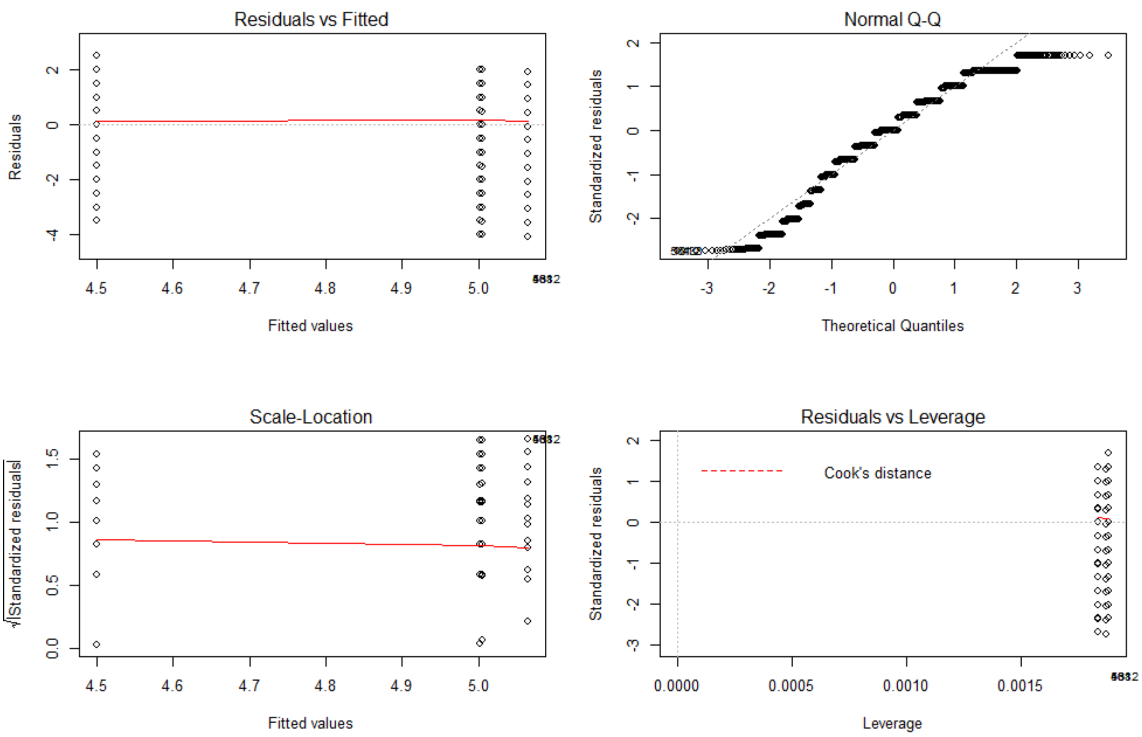
	Perceived trustworthiness		**Use in decision making**	
Condition	***Mean***	***SD***	***Mean***	***SD***
Control	4.20	1.53	5.01	1.47
Ambiguous	3.65	1.51	4.50	1.58
High-agree	4.56	1.45	5.06	1.40
High-data	4.41	1.64	5.00	1.48

Model Diagnostics

Perceived trustworthiness:



Decision Making:



Experimental balance checks

We run balance checks to test whether random assignment successfully balanced demographic background variables across experimental groups, as outlined in the pre-registration. Results are detailed below.

Gender: $\chi^2(3,2151) = 1.59$, $p = .662$

Age: $F(3, 2139) = 1.54$, $p = .203$

Education (treated as categorical): $\chi^2(15,2155) = 8.08$, $p = .920$

Education (treated as continuous): $F(3, 2151) = 0.74$, $p = .526$

Politics: $F(3, 2149) = 2.97$, $p = .031$

Where random assignment did not successfully balance the variables across our experimental groups, we control for these imbalances in our models, as per our pre-registration.

Perceived trustworthiness:

Test of main effect of experimental condition controlling for politics: $F(3, 2146) = 35.84$, $p < .001$, $\eta_p^2 = 0.048$

Decision making:

Test of main effect of experimental condition controlling for politics: $F(3, 2146) = 16.81$, $p < .001$, $\eta_p^2 = 0.023$

Experimental effects remain significant after controlling for political orientation.

Sampling platform check

We ran two-way analysis of variance to test for potential effects of sampling platform on our experimental results. Experimental effects stayed significant controlling for sampling platform. Additionally, we do not find any significant interactions for any of the measures.

Perceived trustworthiness:

Test of main effect of experimental condition controlling for sampling platform: $F(3, 2148) = 36.36$, $p < .001$, $\eta_p^2 = 0.048$

Test of interaction between sampling platform and experimental condition: $F(3, 2145) = 1.16$, $p = .324$

Decision making:

Test of main effect of experimental condition controlling for sampling platform: $F(3, 2148) = 16.85$, $p < .001$, $\eta_p^2 = 0.023$

Test of interaction between sampling platform and experimental condition: $F(3, 2145) = 1.77$, $p = .152$

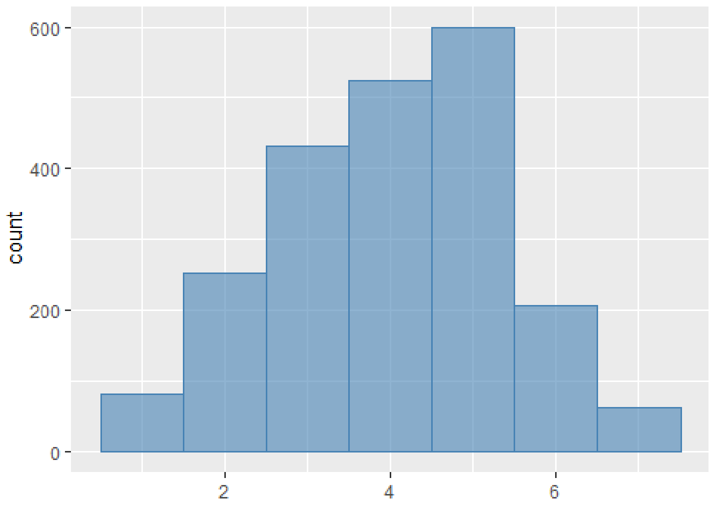
Knowledge measure analysis

Across all three studies, participants were asked how much they thought was currently known about COVID-19. This measure was found to be uninformative due to problems with its design, and hence the reporting was removed from the main paper, as suggested in the review process. Analyses are reported below. A discussion of the measure's design issues can be found in the knowledge measure section of Study 3 in the supplement.

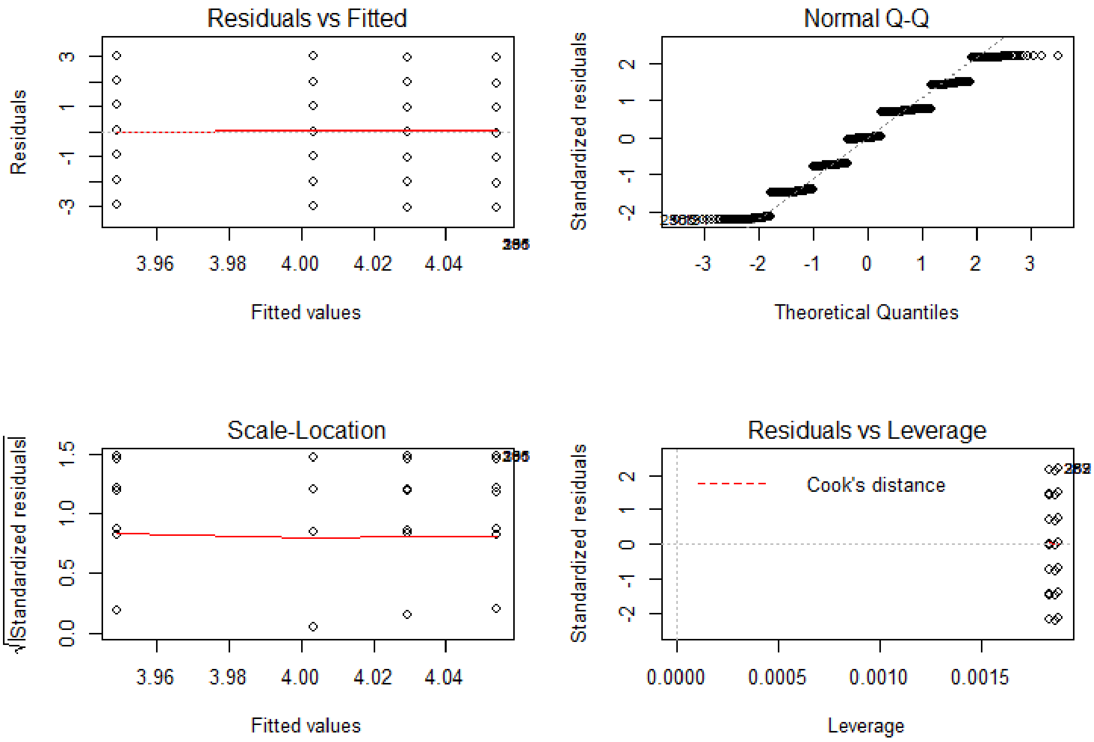
Knowledge item:

How much do you think is currently known about COVID-19? [1-Very little, 7-Very much]

Distribution of knowledge measure:



Model diagnostics:



As in Study 1, we do not find a significant difference between the experimental groups for perceived knowledge ($F(3, 2151) = 0.56$, $p = .641$). Non-parametric robustness testing results were in line with the parametric findings (Kruskal-Wallis $\chi^2 = 1.72$, df = 3, $p = .632$).

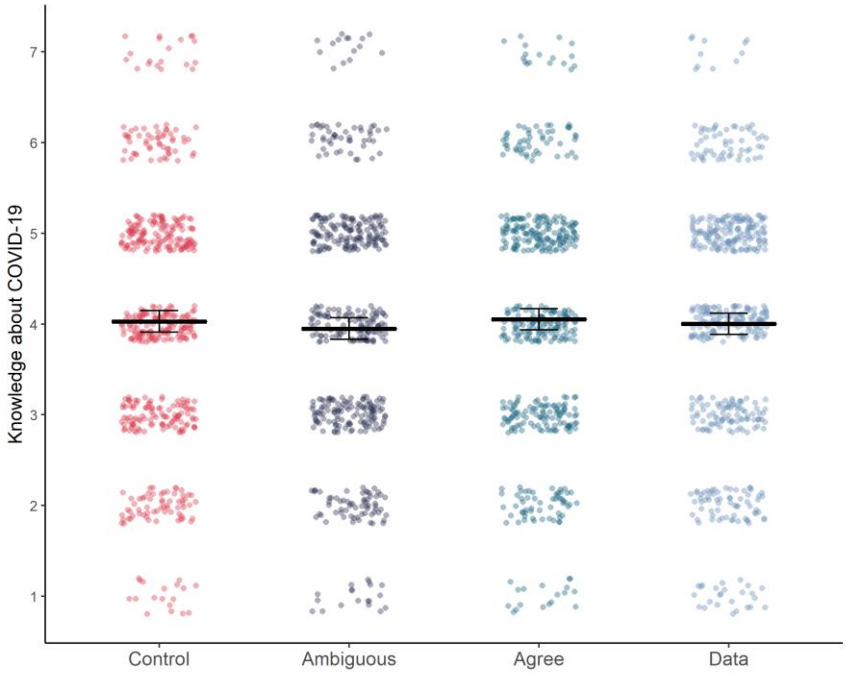


Fig S8. Experimental effects of high quality of evidence (as indicated by agreement between experts and ample availability of data) versus control and ambiguous quality of evidence on knowledge about COVID-19. Error bars denote 95% confidence interval.

Direct effects analysis for SEM modelling contrast selection

Direct effects models* of experimental condition on decision making were run for all contrasts in Study 2, using the *lavaan* package in R (1,000 samples; bootstrapped confidence intervals). Significant contrasts form the basis of the structural equation modelling analysis reported in the main manuscript.

Contrast	Estimate	95% CI
Control - Ambiguous	-0.51	[-0.69, -0.31]
Control – High-agree	0.06	[-0.12, 0.23]
Control – High-data	0	[-0.19, 0.17]
Ambiguous – High-agree	0.57	[0.39, 0.76]
Ambiguous – High-data	0.50	[0.30, 0.68]
High-agree – High-data	-0.06	[-0.23, 0.11]

*While we modelled latent variables for our SEM path analysis reported in the main manuscript, we were not able to include latent variables in our test of the direct effects for contrast selection, as the models in *lavaan* failed to converge for several of the contrasts. We hence base our contrast selection on direct effects modelling which does not include latent variables. Note that a comparison between those direct effects latent variable models that did converge and the respective direct effects models without latent variables revealed differences only in the estimates; however, all contrasts that were significant in the latent variables models were also significant in the models without latent variables.

Exploration of alternative path models

Please refer to the 'alternative path models' section of Study 1 for more background on the alternative models as well as details on their comparison to the main serial model presented in the paper.

In the following, we present results for alternative model 1, i.e. a serial mediation model in which the quality of evidence manipulation influences perceived trustworthiness, which influences perceived uncertainty, and finally decision making; and alternative model 2, i.e. a parallel mediation model in which both uncertainty and trustworthiness jointly affect decision making.

Alternative model 1 (serial mediation) results:

Contrast	Indirect effect		Model fit							
	Estimate	95% CI	χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Ambiguous	-0.22	[-0.32, -0.14]	171.56	13	< .001	0.972	0.107	0.076	19143.78	19213.49
Ambiguous vs. High-agree	0.42	[0.32, 0.52]	178.61	13	< .001	0.971	0.109	0.069	19010.51	19080.10
Ambiguous vs. High-data	0.33	[0.23, 0.44]	191.76	13	< .001	0.970	0.113	0.072	19275.07	19344.78

Indirect effects are significant for all investigated contrasts, however, model fit is generally worse compared to the main model: CFI values are above 0.95, however, lower than for the main model. RMSEA values indicate potential fit issues, as all are above 0.06. SRMR values are just below 0.8 but all substantially higher than for the main model.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 1 (serial mediation):

Contrast	Main serial mediation model in paper		Alternative model 1 (serial mediation)		Difference (Alt1-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Ambiguous	18998.80	19068.51	19143.78	19213.49	144.98	144.98
Ambiguous vs. High-agree	18845.87	18915.46	19010.51	19080.10	164.64	164.64
Ambiguous vs. High-data	19110.00	19179.71	19275.07	19344.78	165.07	165.07

Both AIC and BIC deltas (Alt1 – Main) are positive for all contrasts, suggesting model fit of the main serial model is better compared to the alternative serial model. Furthermore, all observed AIC and BIC differences were >10 providing very strong evidence for the superiority of the main model compared to the serial alternative, following conventional rules of thumb (Burnham and Anderson, 2004; Raftery 1995, 1999) (see Study 1 ‘alternative path models’ section for more details).

Alternative model 2 (parallel mediation) results:

Contrast	**Indirect effect - uncertainty**		**Indirect effect - trust**		**Model fit**							
	Estimate	**95% CI**	**Estimate**	**95% CI**	χ^2	***df***	***p***	**CFI**	**RMSEA**	**SRMR**	**AIC**	**BIC**
Control vs. Ambiguous	0.11	[0.04, 0.20]	-0.4	[-0.57, -0.26]	18.35	11	0.074	0.999	0.025	0.019	18994.57	19074.24
Ambiguous vs. High-agree	-0.11	[-0.25, 0.02]	0.66	[0.49, 0.89]	10.61	11	0.477	1	0	0.008	18846.51	18926.04
Ambiguous vs. High-data	-0.09	[-0.19, 0.01]	0.51	[0.36, 0.68]	23.11	11	0.017	0.998	0.032	0.011	19110.42	19190.08

Only for the Control – Ambiguous contrast did both indirect effect paths emerge as significant. For the Ambiguous – High-agree contrast and the Ambiguous – High-data contrast only the trust path emerged as significant, but not the uncertainty path. This indicates that overall, taken together across all investigated contrasts, the effect of experimental condition on decision making might not be mediated through two separate paths, but that another model (e.g., serial) might be better suited to describe the relationships at play.

Model fit indices (CFI, RMSEA, and SRMR) are comparable to those of the main model presented in the paper.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 2 (parallel mediation):

Contrast	Main serial mediation model in paper		Alternative model 2 (parallel mediation)		Difference (Alt2-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Ambiguous	18998.80	19068.51	18994.57	19074.24	-4.23	5.73
Ambiguous vs. High-agree	18845.87	18915.46	18846.51	18926.04	0.64	10.58
Ambiguous vs. High-data	19110.00	19179.71	19110.42	19190.08	0.42	10.37

BIC deltas (Alt2 – Main) are positive for all contrasts, suggesting model fit of the main serial model is better compared to the alternative parallel model. For two of the three contrasts BIC differences are >10 providing very strong evidence for the superiority of the main model over the parallel model, following conventional rules of thumb (Raftery 1995, 1999). See ‘alternative path models’ section for Study 1 for more details on the employed rules of thumb. For one contrast the BIC difference is between 2 and 6, offering 'positive' evidence that the two models are different in fit and plausibility (Raftery 1995, 1999).

AIC deltas are positive for two of the three contrasts, suggesting model fit of the main serial model is better compared to the alternative parallel model. However, AIC differences are ≤ 2 offering 'substantial support' that the compared models are similar in fit and plausibility (Burnham and Anderson, 2004). For one contrast the AIC delta is negative, suggesting better model fit of the parallel alternative; however, the difference is between 2 and 7, suggesting 'some support' only (Burnham, Anderson, & Huyvaert, 2011).

As indicated in the 'alternative path models' section of Study 1, differences between AIC and BIC can occur in structural equation modelling in some circumstances, given inherent differences in properties and assumptions of the two measures (Vrieze, 2012).

Taken together, the balance of evidence – taking both model fit as well as (non-)significance of indirect effect paths into account - points towards the main serial model being more plausible compared to the parallel alternative.

Thus, overall the evidence suggests that the main model presented in the paper is a better description of the mechanism at play, compared to the investigated alternatives.

Exploratory analysis of additional outcome measures

In addition to the outcome measures reported in the main text, Study 2 collected several other measures, which allowed us to identify the most sensitive and appropriate measures to include in our main confirmatory Study 3. Those measures which showed sever skew/ceiling effects in their distributions, or were thought to be prone to misinterpretation by the participants were consequently dropped from the design and no longer collected in Study 3.

Results for all measures that were collected in Study 2 are reported in the following.

Overview of additional measures collected:

1. How does the information you just saw make you feel? Please indicate this by dragging the slider to select a number from 0 = negative/unhappy to 10 = positive/happy.

2. How easy or difficult do you find this information to understand? [1-Very difficult, 7-Very easy]

3. To what extent do you think the people who are responsible for calculating the COVID-19 case fatality rate mentioned are trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

4. How likely are you to provide the COVID-19 case fatality rate mentioned to a friend if they ask you for it? [1-Not at all likely, 7-Very likely]

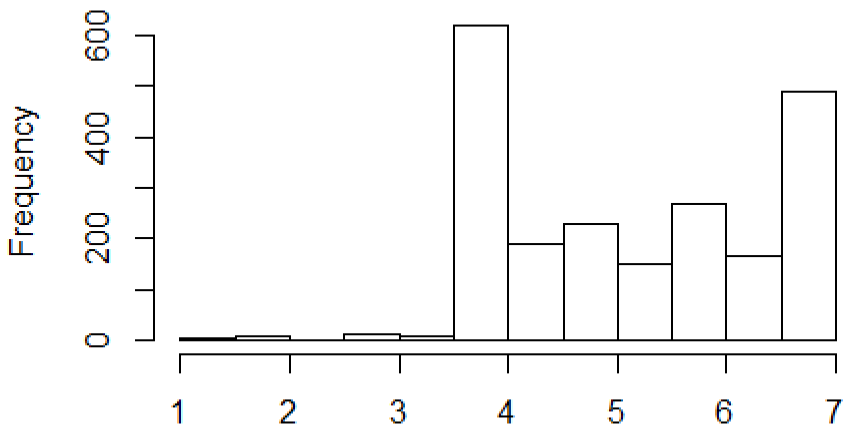
5. How strongly do you agree or disagree with the following statement: *The government should make the mentioned COVID-19 case fatality rate publicly available to everyone on its website about the disease.* [1-Strongly disagree, 7-Strongly agree]

6. Reading the COVID-19 case fatality rate information that we just showed you, how much more or less likely are you to… (1- at lot less likely, 7 – a lot more likely)

- Stay at home (following the government's lock down mandates)
- Physically distance yourself by 2 meters from others when spending time outside
- Self-isolate yourself at home for 14 days if you have recently been around someone who might be infected
- Not touch your face, including your eyes, nose, and mouth, unless your hands are clean
- Cover your nose and mouth when in public

Appraisal:

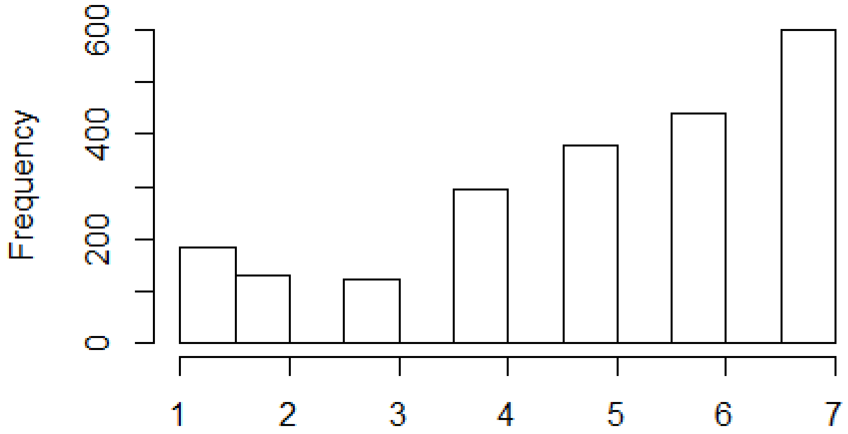
Item 6: Information's effect on compliance with health mandates - The five items of the compliance with public health mandates were combined into an index ($\alpha = 0.92$). Our measure of compliance with public health mandates produced a very non normal distribution with most people reporting to not have changed because of the presented information, and the remaining data showing a skew towards being a lot more likely to comply because of the shown information. Because of this severe non-normal distribution and skew of the data, this measure was subsequently dropped from the study design in Study 3 and not collected.



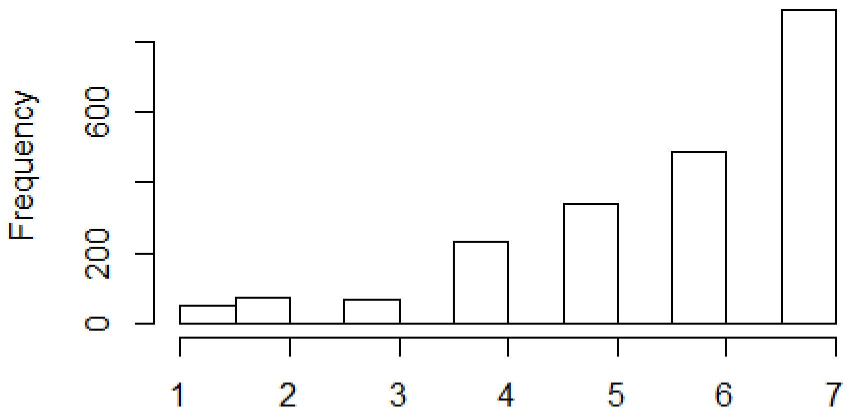
Items 4 (Sharing of information with friends) and 5 (Reporting of information by government): These two measures were intended to capture trust in the information. However, it was deemed that people might also share information that they deem not trustworthy in an effort to make their friends aware of non trustworthy information. Likewise, people could opt to have the government share the information even if they don’t trust it, in an effort to be transparent about low quality data with

people. Apart from these conceptual concerns with the two measures, the distributions were skewed, especially for the government item, hence these two measures were not collected in Study 3 anymore.

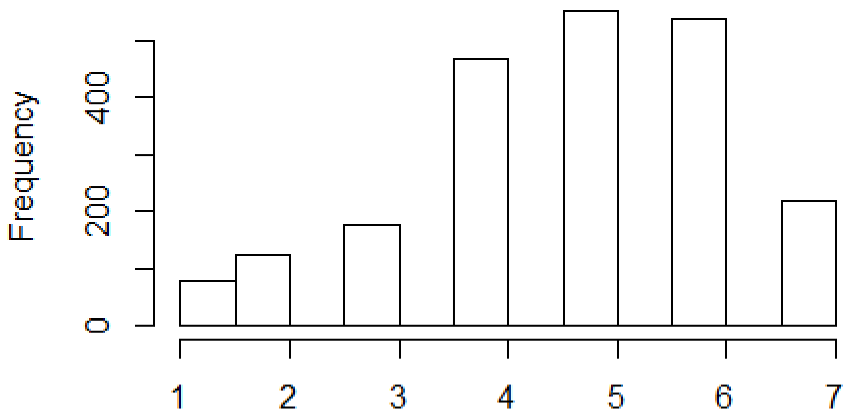
Info sharing with friends:



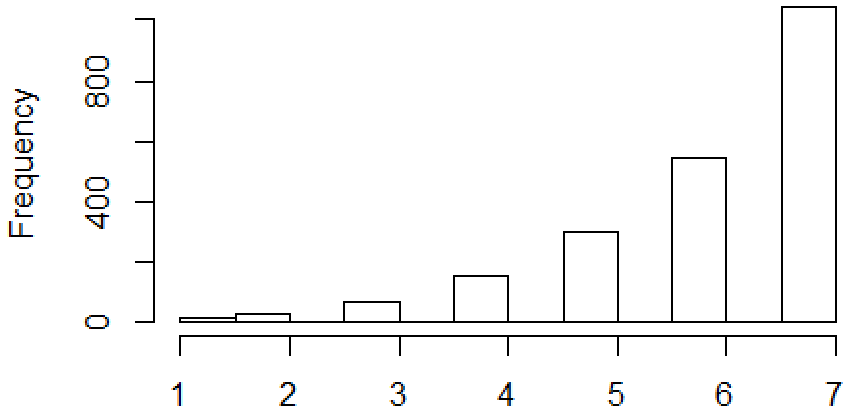
Info sharing by government:



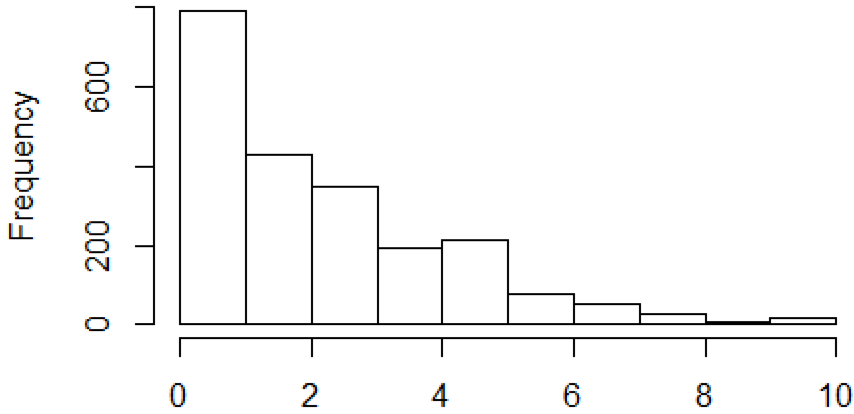
Item 3: Perceived trustworthiness of producers - This measure was kept as an exploratory outcome variable in Study 3.



Item 2: Comprehension - Though comprehension is of interest, we note that this question could not tease apart which aspect the information people were referring to in answering this question: The actual case fatality rate information or rather the quality of evidence information. Due to this potential confound, as well as due to the strong skew in the distribution, this measure was no longer collected in Study 3.



Item 1: Affect - The measure showed severe skew, which would likely inhibit the ability to detect any experimental effects, if present. Hence it was no longer collected in Study 3.



Analysis results:

In the following we report one-way analysis of variance findings for the various outcomes. We note though that non-normality of some of the measures' distributions should be kept in mind.

Item 1: Affect

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition on affect ($F(3, 2149) = 5.29$, $p = .001$, $\eta_p^2 = 0.007$), such that the ambiguous quality of evidence and high quality of evidence due to agreement between experts groups indicated significantly higher positive affect compared to the control group. See visualization of descriptives (group means and confidence intervals) in Figs S9.

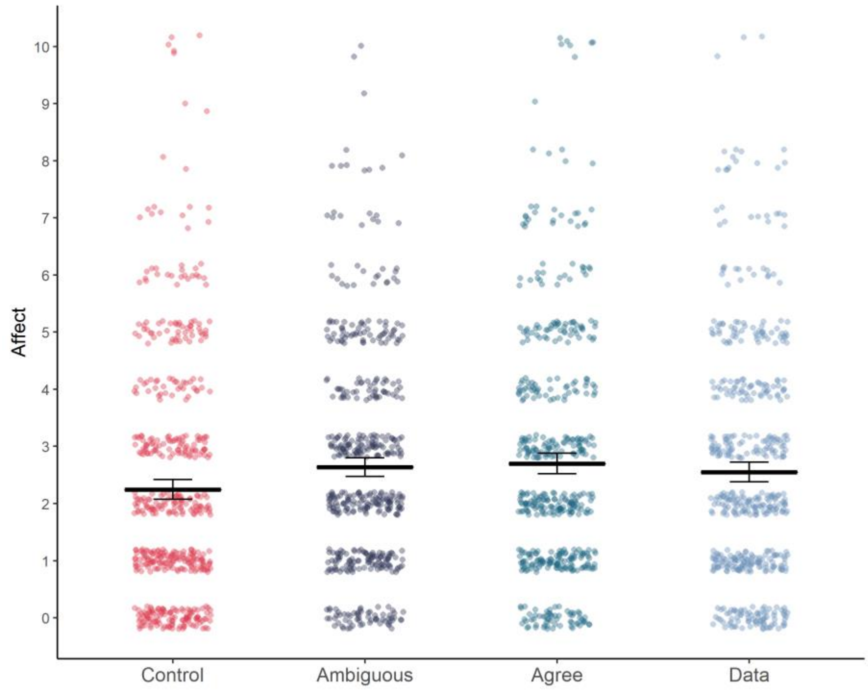


Fig. S9: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on affect. Error bars denote 95% confidence interval.

Item 2: Comprehension

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition on comprehension ($F(3, 2150) = 12.04$, $p < .001$, $\eta_p^2 = 0.017$), such that participants in the ambiguous quality of evidence condition found the information harder to understand compared to participants in the control condition as well as compared to participants in both high quality of evidence conditions. See visualization of descriptives (group means and confidence intervals) in Fig S10.

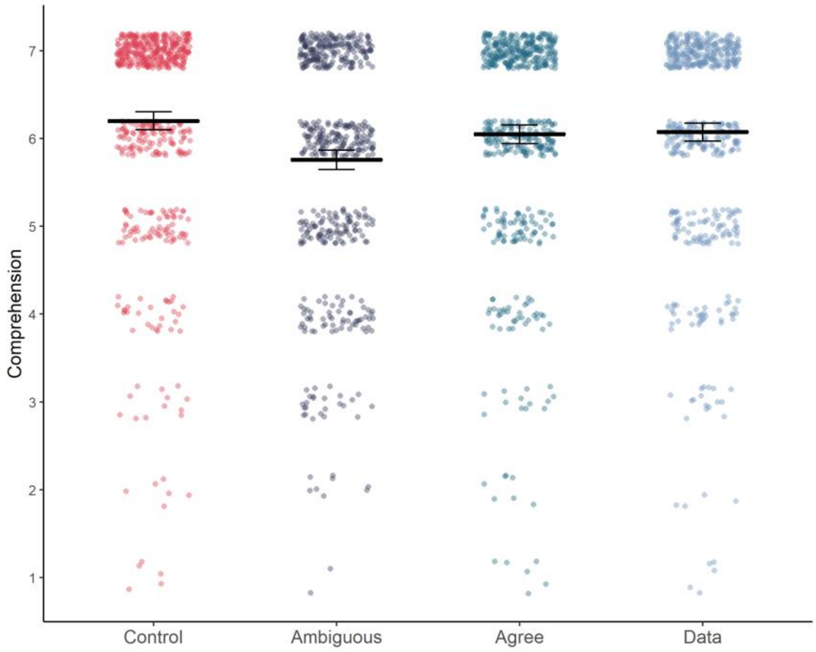


Fig. S10: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on ease of comprehension. Error bars denote 95% confidence interval.

Item 3: Perceived trustworthiness of producers

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition on perceived trustworthiness of the producers of the information ($F(3, 2151) = 13.10$, $p < .001$, $\eta_p^2 = 0.018$), such that participants in both high quality of evidence conditions perceived the producers of the information to be more trustworthy compared to participants in the ambiguous group. Additionally, participants in the high quality of evidence condition as indicated through agreement between experts also indicated higher trustworthiness in the producers compared to control group participants. See visualization of descriptives (group means and confidence intervals) in Fig S11.

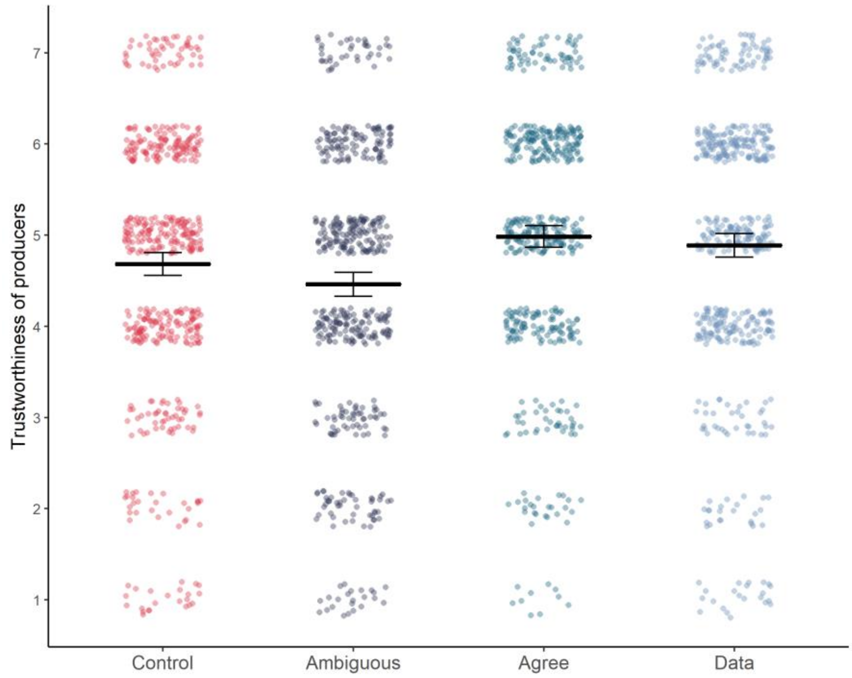


Fig. S11: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on perceived trustworthiness of the producers of the information. Error bars denote 95% confidence interval.

Item 4: Sharing of information with friends

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition on people's inclination to share the case fatality rate information with their friends ($F(3, 2151) = 14.63$, $p < .001$, $\eta_p^2 = 0.020$), such that participants in the two high quality of evidence conditions were significantly more likely to indicate to share the information with friends compared to participants in the ambiguous condition. Additionally, participants in the control group were significantly more likely to share the information with friends than participants in the ambiguous quality of evidence group. See visualization of descriptives (group means and confidence intervals) in Fig S12.

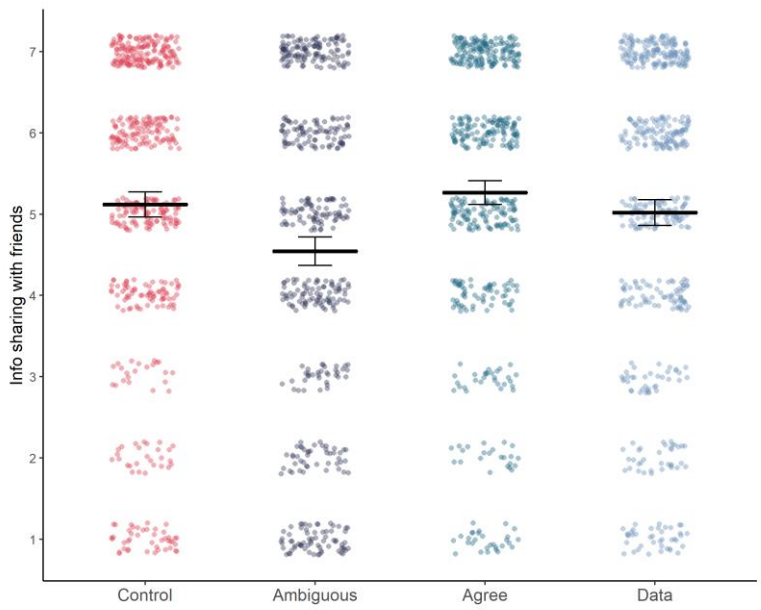


Fig. S12: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on people's intentions to share the case fatality rate information with their friends. Error bars denote 95% confidence interval.

Item 5: Reporting of information by government

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition on people's support of the government making the case fatality rate information publicly available ($F(3, 2149) = 16.05$, $p < .001$, $\eta_p^2 = 0.022$), such that participants in the two high quality of evidence conditions were significantly more likely to indicate that the government should share the information compared to participants in the ambiguous condition. Additionally, participants in the control group were significantly more likely to indicate government sharing of the info than participants in the ambiguous quality of evidence group. See visualization of descriptives (group means and confidence intervals) in Fig S13.

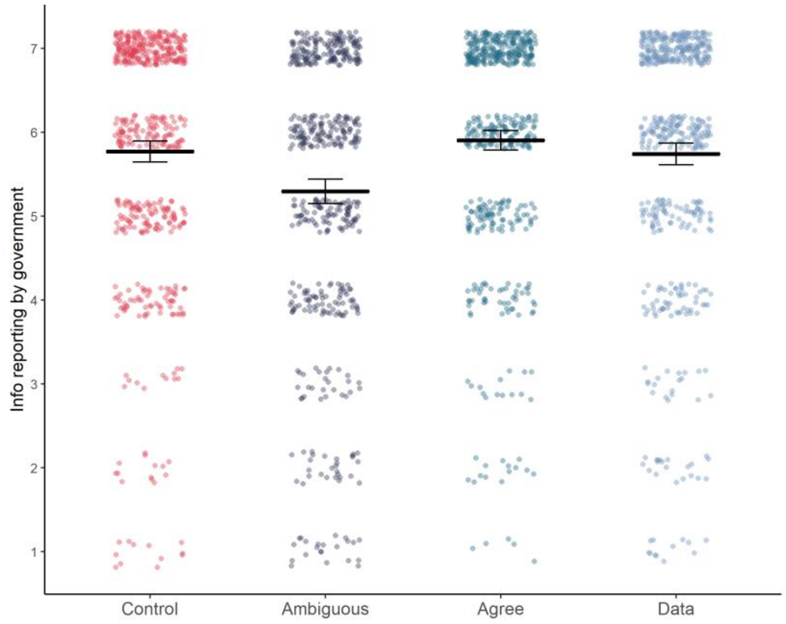


Fig. S13: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on people's support of the government making the case fatality rate information publicly available. Error bars denote 95% confidence interval.

Item 6: Information's effect on compliance with health mandates

One-way analysis of variance using Tukey HSD did not find a significant main effect of experimental condition on people's reported changes in compliance with public health mandates due to the presented information ($F(3, 2150) = 2.48$, $p = .059$). See visualization of descriptives (group means and confidence intervals) in Fig S14.

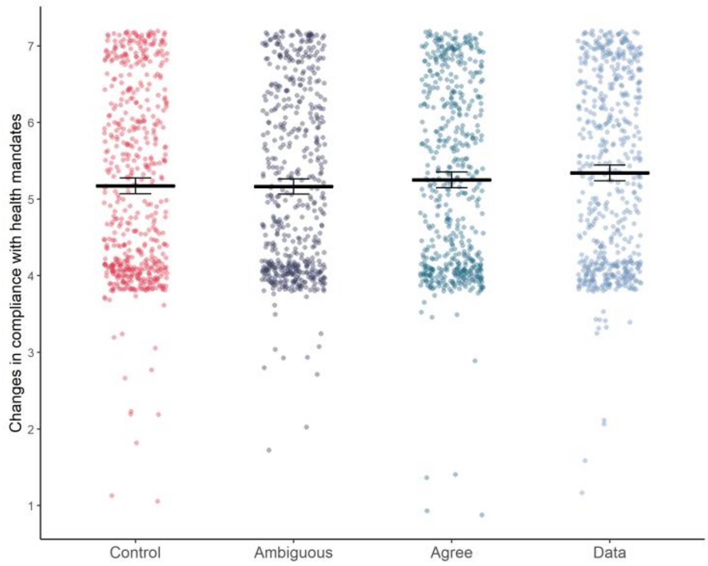


Fig. S14: Experimental effects of high (agreement between experts and ample availability of data) and ambiguous quality of evidence versus control on people's reported changes in compliance with public health mandates. Error bars denote 95% confidence interval.

Study 3

Scenario/Experimental wording

Control group ('Control'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

Ambiguous QoE group ('Ambiguous'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is uncertain. The quality of the evidence could be high or could be low.

<u>Low quality of evidence conditions:</u>

Low QoE group – No explanation ('Low'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is low.

Low QoE group – Disagreement ('Low-disagree'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is low, because there is disagreement between experts.

Low QoE group – Lack of data ('Low-lack'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is low, because there is a lack of data.

<u>High quality of evidence conditions:</u>

High QoE group – No explanation ('High'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is high.

High QoE group – Agreement ('High-agree'):

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is high, because there is a high level of expert agreement.

<u>Robustness check group – to test possible effects of slight wording change in the low QoE conditions between studies 1 (‘uncertain’ wording) and 3 (‘low’ wording) – (Results reported in the section “Robustness check: Wording comparison studies 1 + 3” in the supplement):</u>

Low QoE group – Disagreement – robustness check:

- *Out of every 100 people in the UK who test positive for COVID-19, it is estimated that 14 will die. This is known as the COVID-19 case fatality rate.*

  The quality of the evidence underlying the reported case fatality rate is uncertain, because there is disagreement between experts.

Screenshot of the Oxford COVID-19 Evidence Service website

We used data from the Oxford COVID-19 Evidence Service by the Oxford Centre for Evidence-Based Medicine to provide participants with a current estimate of the case fatality rate for the UK at the time of each study. As the page has been continuously updated, the historic information is not available anymore; we hence include screenshots of the page at the time of Study 3 below (screenshots of the page as of July 6th 2020; retrieved using the wayback machine internet archive, https://web.archive.org/web/20200706085911/https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/). In addition to the case fatality rate, the page also provides information on the uncertainty about the case fatality rate.

CEBM

The Centre for Evidence-Based Medicine develops, promotes and disseminates better evidence for healthcare.

HOME COVID-19 EVIDENCE OPEN EVIDENCE REVIEWS SCHOOL OF EBHC LEADERSHIP
RESOURCES RESEARCH IMPACT OUTREACH BLOG

Global Covid-19 Case Fatality Rates

March 17, 2020

CEBM RESEARCH Oxford COVID-19 Evidence Service

Jason Oke, Carl Heneghan
UPDATED 9th June 2020

Lay Summary by Mandy Payne, Health Watch

This page is updated regularly as new information emerges. It sets out the current Case Fatality Rate (CFR) estimates, the country-specific issues affecting the CFR, and provides a current best estimate of the CFR, and more importantly, the Infection Fatality Rate (IFR).

The IFR estimates the fatality rate in all those with infection: the detected disease (cases) and those with an undetected disease (asymptomatic and not tested group).

Case Fatality Rates:

The total number of cases and the total number of deaths from COVID-19 outbreak data was drawn down (scraped) from https://www.worldometers.info/coronavirus/.

The proportion of deaths to the total numbers of cases was meta-analysed using the R function *metaprop*, using fixed-effect inverse-variance weighting. Estimates from the cruise ship 'Diamond Princess' as well as countries with **fewer than 1000 cases** are excluded from the analysis. (updated 9th April). We present country-level case fatality as a percentage along with 95% confidence intervals in a forest plot. Estimates of heterogeneity and a 95% prediction interval are presented, but a pooled overall estimate is suppressed due to heterogeneity. (understanding data in meta-analysis)

*case fatality rate is the number of reported deaths per number of reported cases (Updated 26th May)

Country	Deaths	Cases	Events per 100 observations	Case Fatality (%)	95%-CI
France	29209	154188		18.94	(18.75 to 19.14)
Belgium	9619	59437		16.18	(15.89 to 16.48)
Italy	33964	235278		14.44	(14.29 to 14.58)
UK	40597	287399		14.13	(14.00 to 14.25)

Last updated: June 09, 2020, 15:21 GMT

Between countries, case Fatality rates vary significantly, and over time, which suggests considerable uncertainty over the exact case fatality rates.

- The number of cases detected by testing will vary considerably by country;
- Selection bias can mean those with severe disease are preferentially tested;
- There may be delays between symptoms onset and deaths which can lead to underestimation of the CFR;
- There may be factors that account for increased death rates such as coinfection, more inadequate healthcare, patient demographics (i.e., older patients might be more prevalent in countries such as Italy);
- There may be increased rates of smoking or comorbidities amongst the fatalities.
- Differences in how deaths are attributed to Coronavirus: dying with the disease (association) is not the same as dying from the disease (causation).

Measures

Mediator index measure

To what extent do you *think* that the COVID-19 case fatality rate mentioned is certain or uncertain?

To what extent do you *feel* that the COVID-19 case fatality rate mentioned is certain or uncertain?

[Answer scale for both:]

- ○ Very certain
- ○ Certain
- ○ Somewhat certain
- ○ Not certain, not uncertain
- ○ Somewhat uncertain
- ○ Uncertain
- ○ Very uncertain

Dependent variables

Perceived trustworthiness index items:

To what extent do you think that the COVID-19 case fatality rate mentioned is accurate? [1-Not accurate at all, 7-Very accurate]

To what extent do you think that the COVID-19 case fatality rate mentioned is reliable? [1-Not reliable at all, 7-Very reliable]

To what extent do you think the COVID-19 case fatality rate mentioned is trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

Decision making index items:

How likely are you to base your own COVID-19 related decisions and behaviours on the mentioned case fatality rate? [1-Not at all likely, 7-Very likely]

To what extent do you think the government should base its decisions and recommendations on how to handle the pandemic on the mentioned COVID-19 case fatality rate? [1-Not at all, 7-Very much]

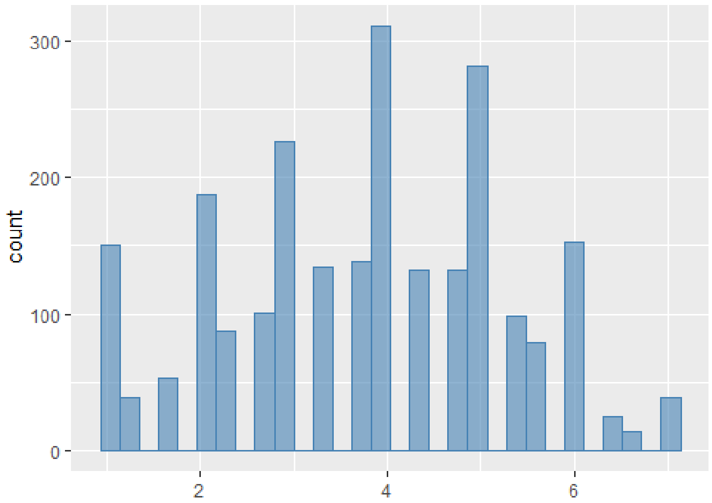
Attention check measure

What was the estimated COVID-19 case fatality rate for the UK that we showed you?

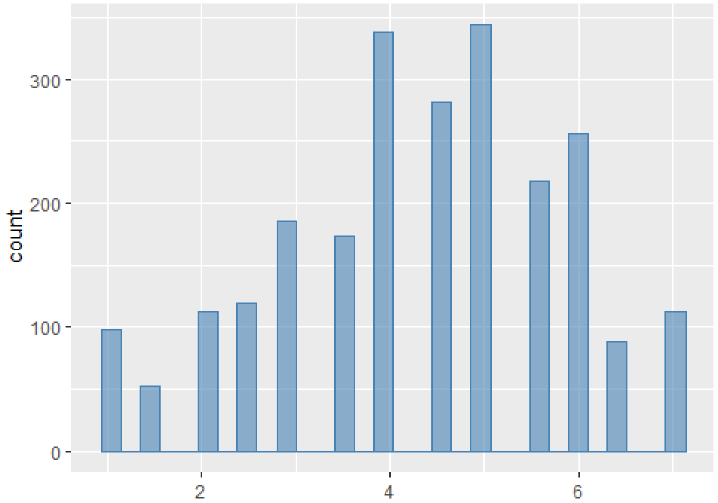
- ◯ 14 out of every 100 people
- ◯ 7 out of every 100 people
- ◯ 18 out of every 100 people
- ◯ 11 out of every 100 people

Distributions of main measures

Perceived trustworthiness:



Decision making:



Demographic composition of study sample

Table S5. Demographic characteristics of the study sample (Study 3).

Variable	**Study 3** (N = 2,392)
Gender, %	
Females	52.38
Males	47.37
Age, *M* (*SD*)	45.16 (16.18)
Education, %	83.19
Political views, *M* (*SD*)	3.73 (1.35)

Note: Due to some participants choosing not to indicate gender, percentages do not total to 100.

*Educational Attainment = *at least Bachelors Degree or equivalent.*

*Political views on spectrum from left wing (or liberal) to right wing (or conservative) on 7-point scale.

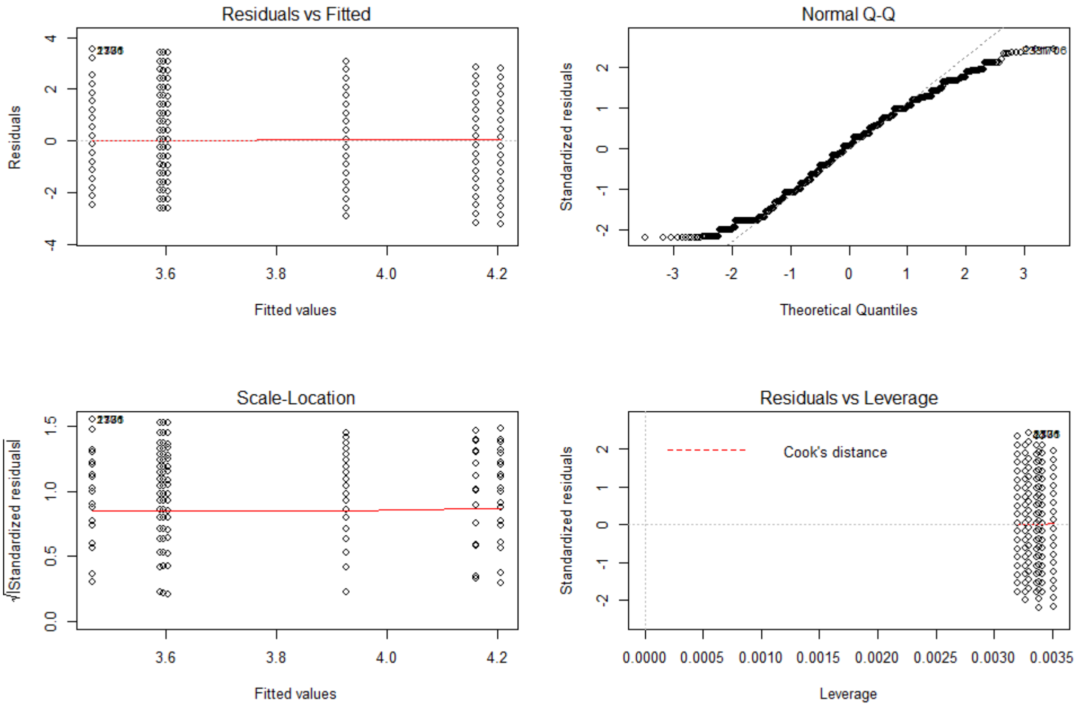
Descriptive statistics for main outcome measures

Table S6. Means and standard deviations per experimental group for perceived trustworthiness and use in decision making.

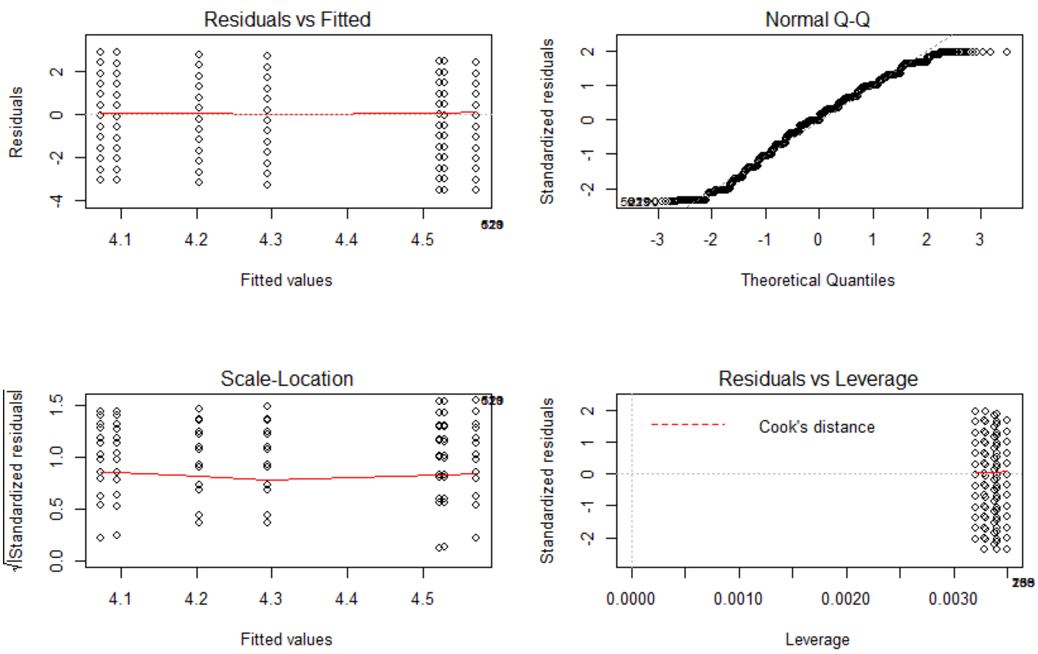
	Perceived trustworthiness		**Use in decision making**	
Condition	***Mean***	***SD***	***Mean***	***SD***
Control	3.93	1.51	4.52	1.56
Ambiguous	3.60	1.46	4.07	1.59
Low	3.47	1.41	4.09	1.50
Low-disagree	3.60	1.39	4.29	1.32
Low-lack	3.59	1.46	4.20	1.51
High	4.16	1.50	4.53	1.47
High-agree	4.21	1.53	4.57	1.50

Model Diagnostics

Perceived trustworthiness:



Decision Making:



Experimental balance checks

We run balance checks to test whether random assignment successfully balanced demographic background variables across experimental groups, as outlined in the pre-registration. Results are detailed below.

Gender: $\chi^2(7,2386) = 2.78$, $p = .905$

Age: $F(7, 2367) = 1.17$, $p = .316$

Education (treated as categorical): $\chi^2(35,2388) = 44.85$, $p = .123$

Education (treated as continuous): $F(7, 2380) = 0.64$, $p = .722$

Politics: $F(7, 2377) = 1.53$, $p = .153$

Sampling platform check

We ran two-way analysis of variance to test for potential effects of sampling platform on our experimental results. Experimental effects stayed significant controlling for sampling platform. Additionally, we do not find any significant interactions for any of the measures.

Perceived trustworthiness:

Test of main effect of experimental condition controlling for sampling platform: $F(6, 2083) = 12.77$, $p < .001$, $\eta_p^2 = 0.035$

Test of interaction between sampling platform and experimental condition: $F(6, 2077) = 1.24$, $p = .285$

Decision making:

Test of main effect of experimental condition controlling for sampling platform: $F(6, 2083) = 6.32$, $p < .001$, $\eta_p^2 = 0.018$

Test of interaction between sampling platform and experimental condition: $F(6, 2077) = 0.65$, $p = .689$

Robustness check: Wording comparison studies 1 + 3

In the context of expert disagreement, we test whether describing the quality of evidence level as 'uncertain' versus as 'low' made a difference on our outcome measures.

'Uncertain' wording: *"The quality of the evidence underlying the reported case fatality rate is uncertain, because there is disagreement between experts".*

'Low' wording: *The quality of the evidence underlying the reported case fatality rate is low, because there is disagreement between experts.*

We do not find a significant difference between the two experimental groups for none of our two main outcome measures.

Perceived trustworthiness: $t = -1.65$, df = 585.01, $p = .099$

Decision Making: $t = -0.63$, df = 576.61, $p = .532$

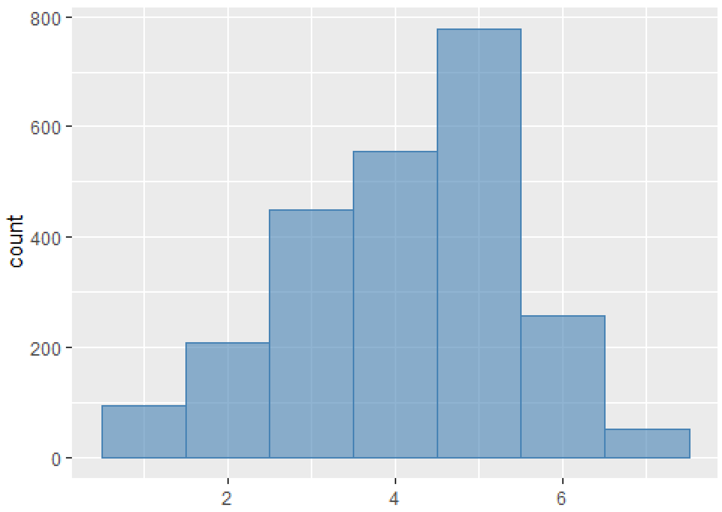
Knowledge measure analysis

Across all three studies, participants were asked how much they thought was currently known about COVID-19. This measure was found to be uninformative due to problems with its design, and hence the reporting was removed from the main paper, as suggested in the review process. Analyses are reported below, together with a discussion of the measure's design issues.

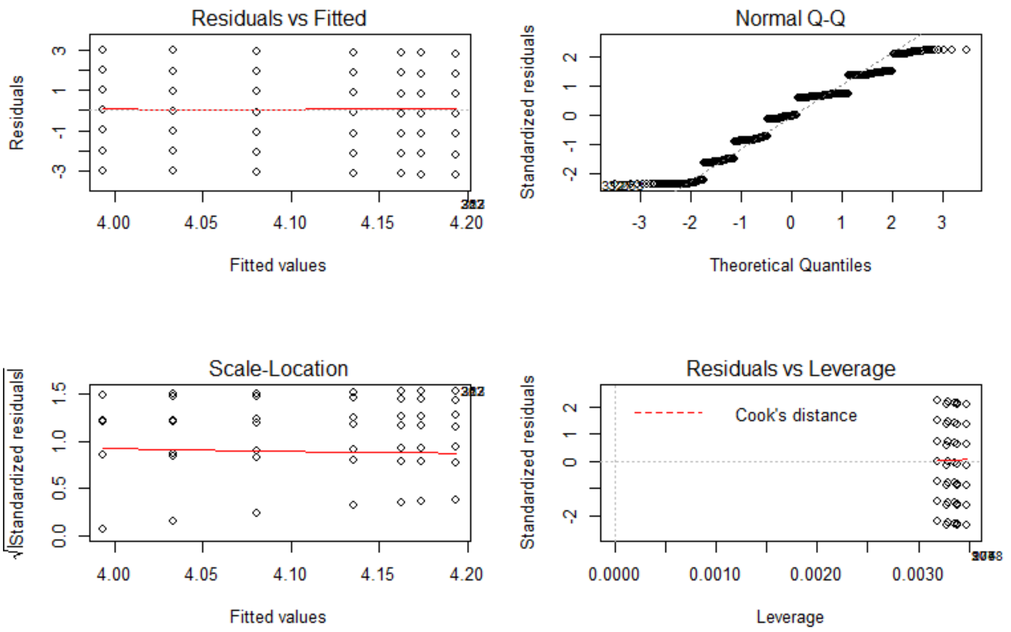
Knowledge item:

How much do you think is currently known about COVID-19? [1-Very little, 7-Very much]

Distribution of knowledge measure:



Model diagnostics:



As in Studies 1 and 2, we did not find a significant effect for perceived knowledge ($F(6, 2091) = 0.96$, $p = .451$). Non-parametric robustness testing results were in line with the parametric findings (Kruskal-Wallis $\chi^2 = 7.17$, df = 6, $p = .305$).

Study 3 thus furthermore corroborates the null-findings on perceived knowledge.

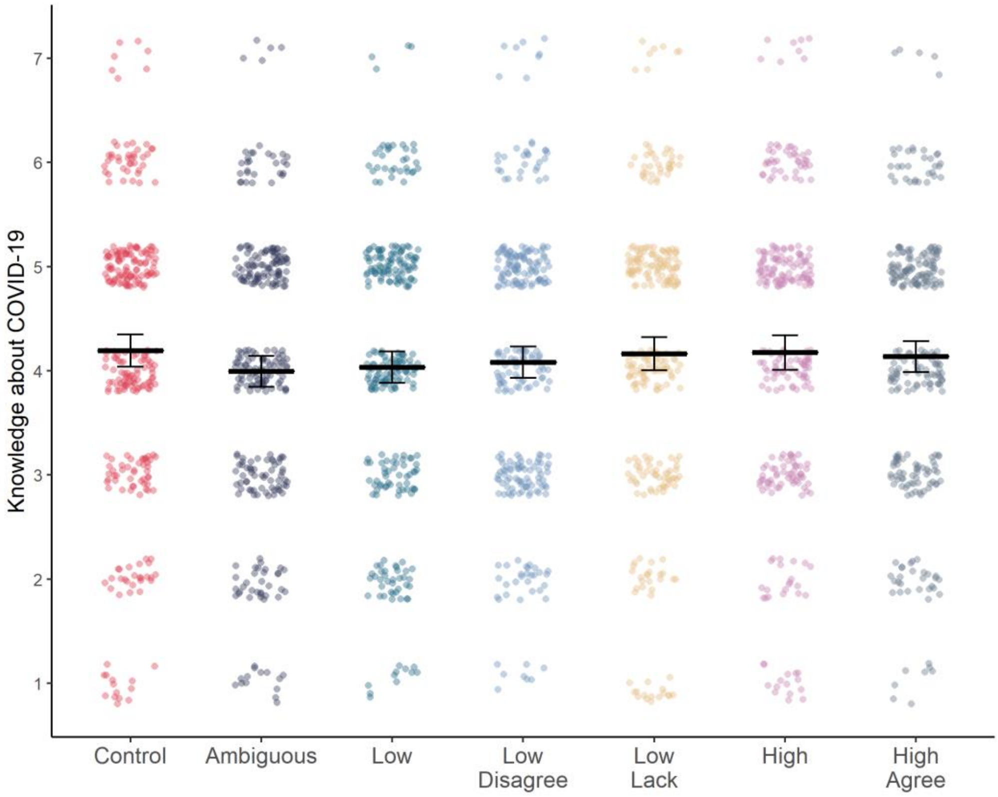


Fig S15. Experimental effects of low quality of evidence (as indicated by disagreement between experts and lack of data, as well as giving no reason) versus high quality of evidence (as indicated by agreement between experts and giving no reason), versus control, versus ambiguous quality of evidence on knowledge about COVID-19. Error bars denote 95% confidence interval.

Discussion:

Interestingly, we did not find any experimental effects on perceived knowledge. Technically, high quality of evidence information provides a stronger knowledge base compared to low quality of evidence information, as low quality signals more potential volatility with regards to the strength of the underlying evidence, while high quality of evidence signals that the provided information is possibly quite robust. Hence, one would expect that perceived knowledge on the subject matter would be increased for high quality of evidence compared to low quality of evidence. However, we do not find such evidence in our data. Although previous research has found that low quality of evidence cues (e.g., conflicting data) leads people to think that “they know less than before” (Koehler & Pennycook, 2019), we find that providing high quality of evidence cues does not lead people to think that “they know more” than before. Koehler and Pennycook (2019) argue that people seem to violate normative scientific inference in the sense that new but conflicting study results can never reduce the state of knowledge. Similarly, it is quite peculiar from a normative perspective that people did not report higher knowledge even though they have clearly obtained more relevant information about the state of the evidence.

It might be that people interpreted our measure ("*How much do you think is currently known about COVID-19?*" very little – very much) in a broader sense, not necessarily drawing a link to the information on the COVID-19 case fatality rate on which they received the quality indicator. Due to this design problem, the knowledge measure is unfortunately of limited informativeness and results can consequently not provide robust insights into the research question. We encourage future research to use more diverse measures of objective and subjective knowledge to further examine possible effects and factors.

Koehler, D. J., & Pennycook, G. (2019). How the public, and scientists, perceive advancement of knowledge from conflicting study results. *Judgment and Decision Making*, *14*(6), 671–682.

Direct effects analysis for SEM modelling contrast selection

Direct effects models* of experimental condition on decision making were run for all contrasts in Study 3, using the *lavaan* package in R (1,000 samples; bootstrapped confidence intervals). Significant contrasts form the basis of the structural equation modelling analysis reported in the main manuscript.

Contrast	**Estimate**	**95% CI**
Control - Ambiguous	-0.45	[-0.70, -0.18]
Control - Low	-0.43	[-0.65, -0.19]
Control - High	0.01	[-0.23, 0.25]
Control – Low-disagree	-0.23	[-0.47, 0.01]
Control – High-agree	0.05	[-0.19, 0.30]
Control – Low-lack	-0.32	[-0.59, -0.09]
Ambiguous - Low	0.02	[-0.21, 0.27]
Ambiguous - High	0.46	[0.23, 0.71]
Ambiguous – Low-disagree	0.22	[-0.02, 0.46]
Ambiguous – High-agree	0.50	[0.24, 0.73]
Ambiguous – Low-lack	0.13	[-0.12, 0.38]
Low - High	0.44	[0.18, 0.67]
Low – Low-disagree	0.20	[-0.03, 0.43]
Low – High-agree	0.48	[0.25, 0.72]
Low – Low-lack	0.11	[-0.14, 0.34]
High – Low-disagree	-0.24	[-0.48, -0.03]
High – High-agree	0.04	[-0.20, 0.27]
High – Low-lack	-0.33	[-0.60, -0.11]
Low-disagree – High-agree	0.28	[0.06, 0.51]
Low-disagree – Low-lack	-0.09	[-0.31, 0.15]
High-agree – Low-lack	-0.37	[-0.60, -0.13]

*While we modelled latent variables for our SEM path analysis reported in the main manuscript, we were not able to include latent variables in our test of the direct effects for contrast selection, as the models in *lavaan* failed to converge for several of the contrasts. We hence base our contrast selection on direct effects modelling which does not include latent variables. Note that a comparison between those direct effects latent variable models that did converge and the respective direct effects models without latent variables revealed differences only in the estimates; however, all contrasts that were significant in the latent variables models were also significant in the models without latent variables.

Exploration of alternative path models

Please refer to the 'alternative path models' section of Study 1 for more background on the alternative models as well as details on their comparison to the main serial model presented in the paper.

In the following, we present results for alternative model 1, i.e. a serial mediation model in which the quality of evidence manipulation influences perceived trustworthiness, which influences perceived uncertainty, and finally decision making; and alternative model 2, i.e. a parallel mediation model in which both uncertainty and trustworthiness jointly affect decision making.

Alternative model 1 (serial mediation) results:

Contrast	Indirect effect		Model fit							
	Estimate	**95% CI**	χ^2	*df*	*p*	**CFI**	**RMSEA**	**SRMR**	**AIC**	**BIC**
Control vs. Ambiguous	-0.17	[-0.31, -0.06]	149.71	18	< .001	0.968	0.109	0.083	12287.79	12362.96
Control vs. Low	-0.23	[-0.37, -0.11]	129.70	18	< .001	0.973	0.101	0.078	12050.99	12125.96
Control vs. Low-lack	-0.16	[-0.26, -0.04]	141.15	18	< .001	0.969	0.107	0.087	11908.21	11982.87
Ambiguous vs. High	0.28	[0.16, 0.42]	183.01	18	< .001	0.958	0.124	0.097	12047.88	12122.48
Ambiguous vs. High-agree	0.29	[0.17, 0.44]	210.91	18	< .001	0.954	0.133	0.096	12135.77	12210.63
Low vs. High	0.34	[0.21, 0.50]	175.04	18	< .001	0.960	0.122	0.090	11813.89	11888.30
Low vs. High-agree	0.35	[0.24, 0.50]	183.29	18	< .001	0.960	0.124	0.090	11885.92	11960.58
High vs. Low-disagree	-0.24	[-0.37, -0.14]	190.48	18	< .001	0.953	0.129	0.094	11596.51	11670.68
High vs. Low-lack	-0.25	[-0.38, -0.14]	189.94	18	< .001	0.954	0.129	0.102	11724.08	11798.16
Low-disagree vs. High-agree	0.26	[0.17, 0.40]	198.26	18	< .001	0.954	0.130	0.092	11667.57	11742.00
High-agree vs. Low-lack	-0.27	[-0.39, -0.16]	183.47	18	< .001	0.959	0.125	0.100	11761.50	11835.85

Indirect effects are significant for all investigated contrasts, however, model fit is generally worse compared to the main model. CFI values are above 0.95, however, lower than for the main model. RMSEA values indicate potential fit issues, as all are substantially above 0.06. The same for SRMR values which are above 0.08 for all but one contrast.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 1 (serial mediation):

Contrast	Main serial mediation model in paper		Alternative model 1 (serial mediation)		Difference (Alt1-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Ambiguous	12177.26	12252.43	12287.79	12362.96	110.53	110.53
Control vs. Low	11944.12	12019.09	12050.99	12125.96	106.87	106.87
Control vs. Low-lack	11800.02	11874.68	11908.21	11982.87	108.19	108.19
Ambiguous vs. High	11889.21	11963.81	12047.88	12122.48	158.67	158.67
Ambiguous vs. High-agree	11973.94	12048.80	12135.77	12210.63	161.83	161.83
Low vs. High	11660.93	11735.33	11813.89	11888.30	152.96	152.97
Low vs. High-agree	11732.32	11806.99	11885.92	11960.58	153.60	153.59
High vs. Low-disagree	11442.74	11516.91	11596.51	11670.68	153.77	153.77
High vs. Low-lack	11564.23	11638.32	11724.08	11798.16	159.85	159.84
Low-disagree vs. High-agree	11510.92	11585.36	11667.57	11742.00	156.65	156.64
High-agree vs. Low-lack	11600.24	11674.59	11761.50	11835.85	161.26	161.26

Both AIC and BIC deltas (Alt1 – Main) are positive for all contrasts, suggesting model fit of the main serial model is better compared to the alternative serial model. Furthermore, all observed AIC and

BIC differences are >10 providing very strong evidence for the superiority of the main model compared to the serial alternative, following conventional rules of thumb (Burnham and Anderson, 2004; Raftery 1995, 1999) (see Study 1 ‘alternative path models’ section for more details).

Alternative model 2 (parallel mediation) results:

Contrast	Indirect effect - uncertainty		Indirect effect - trust		Model fit							
	Esti-mate	95% CI	Esti-mate	95% CI	χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Ambiguous	-0.10	[-0.23, -0.01]	-0.21	[-0.40, -0.07]	29.92	16	0.018	0.997	0.038	0.016	12172.00	12256.01
Control vs. Low	-0.08	[-0.21, -0.01]	-0.28	[-0.46, -0.13]	18.15	16	0.315	0.999	0.015	0.011	11943.44	12027.23
Control vs. Low-lack	-0.11	[-0.21, -0.02]	-0.18	[-0.33, -0.04]	21.01	16	0.178	0.999	0.023	0.012	11792.07	11875.52
Ambiguous vs. High	0.11	[-0.03, 0.29]	0.36	[0.20, 0.57]	13.98	16	0.600	1	0	0.008	11882.84	11966.23
Ambiguous vs. High-agree	0.04	[-0.10, 0.21]	0.42	[0.23, 0.61]	42.86	16	< .001	0.994	0.053	0.014	11971.72	12055.39
Low vs. High	0.09	[-0.07, 0.25]	0.45	[0.28, 0.65]	17.99	16	0.324	0.999	0.015	0.011	11660.85	11744.01
Low vs. High-agree	0.02	[-0.11, 0.17]	0.50	[0.34, 0.70]	28.64	16	0.027	0.997	0.036	0.010	11735.26	11818.71
High vs. Low-disagree	-0.07	[-0.25, 0.07]	-0.32	[-0.47, -0.17]	21.59	16	0.157	0.998	0.025	0.016	11431.61	11514.51
High vs. Low-lack	-0.11	[-0.26, 0.03]	-0.33	[-0.50, -0.19]	19.17	16	0.260	0.999	0.019	0.014	11557.30	11640.10
Low-disagree vs. High-agree	-0.01	[-0.15, 0.14]	0.38	[0.23, 0.55]	29.86	16	0.019	0.996	0.038	0.014	11503.16	11586.35
High-agree vs. Low-lack	-0.04	[-0.18, 0.08]	-0.39	[-0.55, -0.24]	15.74	16	0.471	1	0	0.010	11597.77	11680.86

While the indirect effect path through trust is significant for all contrasts, the indirect effect path through uncertainty is non-significant for the majority of contrasts (i.e., 8 out of 11). This suggests that overall, taken together across all investigated contrasts, the effect of experimental condition on decision making might not be mediated through two separate paths (i.e., not mediated jointly through trust and uncertainty), but that another model (e.g., serial) might be better suited to describe the relationships at play.

Model fit indices (CFI, RMSEA, and SRMR) are comparable to those of the main model presented in the paper.

Akaike and Bayesian Information Criterion model comparison between main model and alternative model 2 (parallel mediation):

Contrast	Main serial mediation model in paper		Alternative model 2 (parallel mediation)		Difference (Alt2-Main)	
	AIC	BIC	AIC	BIC	AIC	BIC
Control vs. Ambiguous	12177.26	12252.43	12172.00	12256.01	-5.26	3.58
Control vs. Low	11944.12	12019.09	11943.44	12027.23	-0.68	8.14
Control vs. Low-lack	11800.02	11874.68	11792.07	11875.52	-7.95	0.84
Ambiguous vs. High	11889.21	11963.81	11882.84	11966.23	-6.37	2.42
Ambiguous vs. High-agree	11973.94	12048.80	11971.72	12055.39	-2.22	6.59
Low vs. High	11660.93	11735.33	11660.85	11744.01	-0.08	8.68
Low vs. High-agree	11732.32	11806.99	11735.26	11818.71	2.94	11.72
High vs. Low-disagree	11442.74	11516.91	11431.61	11514.51	-11.13	-2.40
High vs. Low-lack	11564.23	11638.32	11557.30	11640.10	-6.93	1.78
Low-disagree vs. High-agree	11510.92	11585.36	11503.16	11586.35	-7.76	0.99
High-agree vs. Low-lack	11600.24	11674.59	11597.77	11680.86	-2.47	6.27

We observe some differences between the various contrasts in conclusions drawn from AIC and BIC comparison, as can occur in structural equation modelling given inherent differences in properties and assumptions of the two measures (Vrieze, 2012).

The majority of BIC differences (10 out of 11) are positive, suggesting that model fit of the main serial model is better compared to the alternative parallel model. For one of the ten contrasts the BIC difference is >10 providing very strong evidence for the superiority of the main model over the parallel model (Raftery 1995, 1999). For four of the ten contrasts BIC difference is between 6 and 10 offering 'strong' evidence. For two of the ten contrasts BIC difference is between 2 and 6 offering 'positive' evidence. For three of the ten contrasts BIC difference is between 0 and 2 offering 'weak' evidence. For one contrast the BIC difference is negative (between 2 and 6), offering 'positive' evidence for better fit of the parallel model for this contrast.

AIC differences for the majority of contrasts (10 out of 11) are negative, suggesting that model fit of the parallel model is better compared to the main serial model. However, for half of the these (five out of ten) absolute AIC difference is between 2 and 7, suggesting only 'some support' (Burnham, Anderson, & Huyvaert, 2011). For two of the ten contrasts, absolute AIC difference is ≤ 2 providing 'substantial support' that the two models might not be distinguishable in fit (Burnham and Anderson, 2004). Only one contrast shows an absolute AIC difference of >10 (suggesting 'essentially no support' that the two models are equivalent for this contrast). Finally, one contrast shows a positive AIC difference, suggesting better fit of the main serial model; noting that the difference is between 2 and 7, i.e., suggesting 'some support' (Burnham, Anderson, & Huyvaert, 2011).

Taken together, the balance of evidence – taking both model fit as well as (non-)significance of indirect effect paths into account - points towards the main serial model being more plausible compared to the parallel alternative.

Thus, overall the evidence suggests that the main model presented in the paper is a better description of the mechanism at play, compared to the investigated alternatives. However, we note these as tentative conclusions and encourage more research – employing confirmatory methods – to examine the underlying relationships in more depth.

Exploratory analysis of additional outcome measure

We had included an exploratory item that explored experimental effects on perceived trustworthiness of the producers of the presented information with the measure:

To what extent do you think the people who are responsible for calculating the COVID-19 case fatality rate mentioned are trustworthy? [1-Not trustworthy at all, 7-Very trustworthy]

One-way analysis of variance using Tukey HSD found a significant main effect of experimental condition ($F(6, 2088) = 4.21$, $p < .001$, $\eta_p^2 = 0.012$), such that participants in the low quality of evidence condition that presented no reason for the quality level, perceived the trustworthiness of the producers of the information to be significantly lower compared to both high quality of evidence conditions. In addition, participants in the high quality of evidence group as indicated through expert agreement indicated significantly higher trustworthiness of the producers compared to participants in the low quality of evidence condition as indicated by expert disagreement. No other contrasts reached significance.

See visualization of descriptives (group means and confidence intervals) in Figs S16.

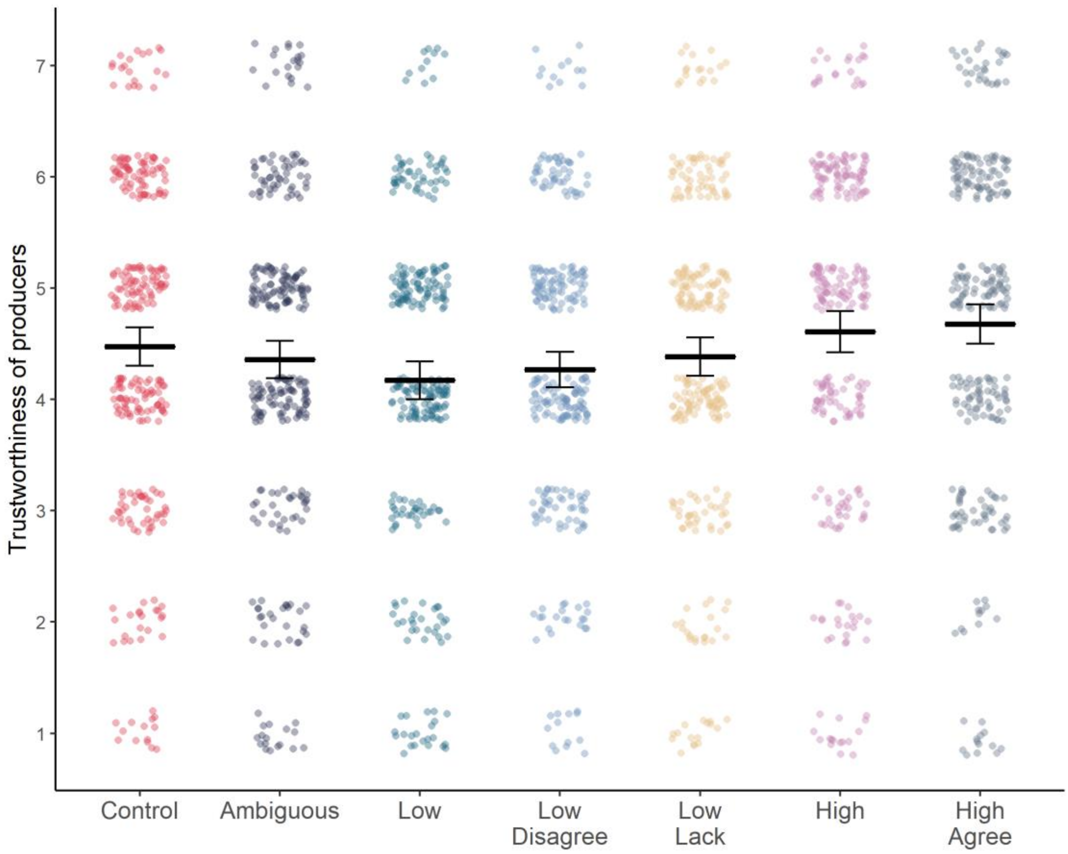


Fig. S16: Experimental effects of high (agreement and no reason), low (disagreement, lack of data, and no reason), and ambiguous quality of evidence versus control on perceived trustworthiness of the producers of the information. Error bars denote 95% confidence interval.