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

Claudia R. Schneider*   Alexandra L. J. Freeman†   David Spiegelhalter‡

Sander van der Linden§

Abstract

Large-scale societal issues such as public health crises highlight the need to communicate scientific information, which is often uncertain, accurately to the public and policy makers. The challenge is to communicate the inherent scientific uncertainty — especially about the underlying quality of the evidence — whilst supporting informed decision making. Little is known about the effects that such scientific uncertainty has on people's judgments of the information. In three experimental studies (total *N*=6,489), we investigate the influence of scientific uncertainty about the quality of the evidence on people's perceived trustworthiness of the information and decision making. We compare the provision of high, low, and ambiguous quality-of-evidence indicators against providing no such cues. Results show an asymmetric relationship: people react more strongly to cues of low quality of evidence than they do to high quality of evidence compared to no cue. While responses to a cue of high quality of evidence are not significantly different from no cue; a cue of low or uncertain quality of evidence is accompanied by lower perceived trustworthiness and lower use of the information in decision making. Cues of uncertain quality of evidence have a similar effect to those of low quality. These effects do not change with the addition of a reason for the indicated quality level. Our findings shed light on the effects of the communication of scientific uncertainty on judgment and decision making, and provide insights for evidence-based communications and informed decision making for policy makers and the public.

Keywords: scientific uncertainty, quality of evidence, science communication, trustworthiness, decision making.

*Winton Centre for Risk and Evidence Communication, and Department of Psychology, University of Cambridge. E-mail: cs2025@cam.ac.uk. https://orcid.org/0000-0002-6612-5186.

†Winton Centre for Risk and Evidence Communication, University of Cambridge. https://orcid.org/0000-0002-4115-161X.

‡Winton Centre for Risk and Evidence Communication, University of Cambridge. https://orcid.org/0000-0001-9350-6745.

§Winton Centre for Risk and Evidence Communication, and Department of Psychology, University of Cambridge. https://orcid.org/0000-0002-0269-1744.

1   Introduction

Science is inherently uncertain, and in a public health context, when dealing for instance with new diseases in a crisis situation, information can be especially uncertain. Yet, decisions need to be made, despite the uncertainty. Science communication, particularly non-persuasive communication in this kind of context, therefore faces the challenge of expressing scientific uncertainty whilst supporting informed decision making (Blastland et al., 2020; Fischhoff & Davis, 2014). Uncertainty, for instance, around an estimate of an effect – 'direct' uncertainty (van der Bles et al., 2019) — is known to play a role in how people perceive information and use it in their decision making (Gustafson & Rice, 2020). The way in which direct uncertainty around a number is communicated, such as verbally or numerically for instance, can lead to different interpretations of the information, and much research has shed light on these relationships (Dhami & Mandel, 2021; Kause et al., 2021; Mandel et al., 2021; Mandel & Irwin, 2021; van der Bles et al., 2020). Additionally, research has shown that the format in which such direct uncertainty (e.g., probabilistic predictions) is communicated can also affect the perceived credibility of the communicator (Collins & Mandel, 2019; Jenkins et al., 2017; Jenkins et al., 2018; Jenkins & Harris, 2021). Although some literature points to a potential adverse effect of presentation format on aspects such as understanding and credibility, disclosure of direct uncertainty around numeric estimates and predictions more generally, does not necessarily seem to decrease trust in the numbers or in the source of the message (Gustafson & Rice, 2019, 2020; Hendriks & Jucks, 2020; Howe et al., 2019; Kreps & Kriner, 2020; van der Bles et al., 2020) and can even increase trust in some contexts (Joslyn & Demnitz, 2021; Joslyn & Leclerc, 2016).

Despite the sizeable body of research and the insights gained on the effects of communicating direct, quantified uncertainties, little is known about the effects of communicating the deeper, more subjective, and not easily quantifiable elements of uncertainty in scientific information. This makes up the 'indirect' uncertainty around a claim or estimate (van der Bles et al., 2019), such as the underlying quality of evidence (Weiss, 2003). Quality of evidence is often broadly defined as "the extent to which confidence in an estimate of the effect is adequate to support recommendations" (Guyatt, Oxman, Vist, Kunz, Falck-Ytter, & Schünemann, 2008, p. 995). Unlike quantifiable uncertainties, which can be communicated in the form of a numeric or confidence interval, it is a more subjective and less easily quantified measure of how much the estimate could be affected by known and unknown factors and biases (e.g., lack of data, disagreement between experts, imprecision, confounding variables).

In a recent article, Blastland et al. (2020) present five rules for effective evidence communication, one of which is to clearly state "evidence quality" (as well as direct uncertainties). Fischhoff and Davis (2014) argue that such uncertainty information needs to be summarized in a useful form for decision makers. In fact, there is a widespread use of qualitative scales for quality-of-evidence communication, mostly in the form of short descriptive labels (e.g., 'high' or 'low' quality of evidence), sometimes accompanied by or symbolised by icons, such as by GRADE (Guyatt, Oxman, Vist, Kunz, Falck-Ytter, Alonso-Coello, et al., 2008), the Education Endowment Foundation (EFF, 2020), SAGE documents (UK Government, 2020), the Admiralty Code in military operations and intelligence analysis (Irwin & Mandel, 2019), or the Intergovernmental Panel on Climate Change reports (IPCC, 2014; Mastrandrea et al., 2010). However, systematic evaluations of the effects of communicating quality of evidence on the audiences are sparse or largely non-existent.

It is conceivable that a cue that a particular claim is based on underlying evidence of high quality may act as a signal that such a claim can be trusted. It signals that potential gains can be taken from knowing the information and using the information in subsequent judgments and decision making. Conversely, it is conceivable that seeing a flag of low quality of evidence may act as a warning sign, making people less likely to use the information. Recent work on the effects of low and high quality evidence underlying the effectiveness of a public health measure supports this proposition; participants who were presented with a low quality-of-evidence indicator for the effectiveness of eye protection against COVID-19 perceived the evidence as less trustworthy, rated the intervention as less effective and in one of two studies indicated lower behavioural intentions to wear eye protection (Schneider et al., 2021). What is more, in a second study which compared high and low quality-of-evidence information to a control group which did not provide a quality indicator, distance from the control group in terms of ratings of perceived trustworthiness and effectiveness differed for the high versus the low quality-of-evidence groups: while perceived trustworthiness and effectiveness were significantly lower in the low quality-of-evidence group compared to the control group, there was no significant difference between the high quality-of-evidence group and the control group (Schneider et al., 2021). These results provide some initial insights that seem to suggest that people may react more strongly to indications of low quality of evidence compared to high quality of evidence.

Another important question is how people make judgments in response to information of explicitly uncertain quality. For example, what if — in a decision context — people are told explicitly that it is not known whether the quality level of the evidence is high or low (i.e., when the quality of evidence is ambiguous). Findings on ambiguity aversion suggest that "the subjective experience of missing information" (Frisch & Baron, 1988, p. 152) would lead people to want to avoid uncertainty (Fox & Tversky, 1995; Tversky & Kahneman, 1974). Often referred to as "one of the most robust phenomena in the decision-making literature" (Keren & Gerritsen, 1999, p. 1), research on ambiguity aversion has shown that people prefer the clear over the vague or certainty over uncertainty across a range of contexts

851

(Camerer & Weber, 1992; Keren & Gerritsen, 1999). This is also an important practical question as it can be hard to assess quality of evidence comprehensively, especially in crises — such as during the COVID-19 pandemic — when urgent information turnaround and knowledge production are paramount. For example, policy makers and communicators might have an estimate of the case fatality rate at hand, but it might not be clear what the quality level of the underlying evidence is.

Our research thus asks the question of how people incorporate information on the quality of the underlying evidence into their judgments and decision making. Do they disregard quality when it is not certain and treat it as if the quality indicator did not exist? Do they assume the best, i.e., do they assume high quality in the absence of clarity? Or do they assume the worst, i.e., put it on par with low quality-of-evidence information? And how does disclosure of quality-of-evidence information, be it high, low, or ambiguous, compare to non-disclosure, namely, cases in which the quality of evidence is completely omitted and not even mentioned as a concept? Where do people's reactions in terms of trust and decision making sit in those cases compared to when they are explicitly told that the quality of the underlying evidence is high, low, or ambiguous?

Another question of interest pertains to the mechanism that underlies possible shifts in perceived trustworthiness and decision making when people are presented with varying levels of quality-of-evidence information. Recent research suggests that conflicting studies and data (an aspect of quality of evidence) may trigger a sense of uncertainty and a feeling that no new knowledge has been gained, even though more data on the matter has accumulated (Koehler & Pennycook, 2019). Similarly, because consensus among experts is an influential judgment cue for people in many decision contexts (van der Linden et al., 2015, 2019), disagreement among experts can also elicit an indirect sense of uncertainty which can negatively influence trust and decision making (Aklin & Urpelainen, 2014; Gustafson & Rice, 2020). Based on these indications from the literature with regards to the role of perceived uncertainty, we suggest that varying levels of quality of evidence could influence people's perception of the uncertainty or certainty of the presented information, which could affect trust and decision making. In addition, recent empirical findings suggest that communicating quality of evidence can influence behavioural indicators through varying perceptions of trustworthiness (Schneider et al., 2021). Taken together, it is thus conceivable that the provision of quality-of-evidence cues can influence people's perceived uncertainty or certainty of the presented information, which could then account for changes in perceived trustworthiness and, in turn, decision making.

In short, despite some initial findings, much is still unknown about how people make judgments and decisions when confronted with cues about the underlying quality of the evidence — especially the question of how people react to ambiguous quality-of-evidence information — and what psychological mechanisms might explain such judgments, from quality perceptions through to decision making. The current research program seeks to remedy these gaps. We address these issues systematically, in the public health domain

852

— using COVID-19 as a real-life context. With well over a million deaths worldwide (ECDC, 2020), there has perhaps never been so much emphasis on the need to "follow the science" and weigh up evidence. However, both the public and policy makers must make overt and wide-reaching decisions based on evidence that is of variable, uncertain, and often low, quality. It is thus crucial to understand how information about the quality of underlying evidence — or a lack of information about it — influences people's appraisal of the evidence, their opinion of its trustworthiness, and behaviour as a result of it.

Accordingly, in three large-scale experimental studies, we explore how people react to high versus low versus ambiguous versus no quality-of-evidence information in the context of a COVID-19 public health message. We also investigate a potential psychological mechanism that could explain people's reactions to quality-of-evidence cues. Study 1 assesses the effects of providing people with low quality-of-evidence information (compared to giving no quality cue, i.e., no mentioning of quality of evidence, not even as a concept). Study 2 investigates effects for high and ambiguous quality-of-evidence information. Study 3 constitutes our main confirmatory study which replicates and extends findings from Studies 1 and 2. Here we test effects of disclosure of low, high, and ambiguous quality of evidence versus non-disclosure against each other. We investigate effects on (a) perceived uncertainty of the provided information, (b) trustworthiness of the information, (c) and downstream decision making.

Based on the literature presented above, we hypothesized that (a) effects (distance from a control group that does not mention quality of evidence) for low quality of evidence are stronger compared to effects for high quality of evidence, and (b) we expected that people are averse to ambiguous or uncertain quality-of-evidence information (compared to more certain information on quality). We additionally test whether the effect of quality-of-evidence information on perceived trustworthiness could be mediated by perceptions of increased (or decreased) uncertainty and whether in turn perceived trustworthiness could mediate the effect on decision making. We furthermore test two different explanations for classifying evidence as either high or low quality (quality variations stemming from expert agreement or disagreement on the UK's COVID-19 case fatality rate and quality variations due to high or low availability of data underlying the calculated case fatality rate) (Studies 1–3) and compare effects of providing an explicit explanation for the level of quality with giving people only the quality level indicator on its own, without further explanation (Study 3).

1.1 Overview of Studies

We ran three studies to investigate the effects of quality-of-evidence communication on perceived trustworthiness and decision making. Our studies are situated in a public health context, providing participants with the UK's estimated COVID-19 case fatality rate alongside varying (or absent) quality-of-evidence information. At the time of data collection for each study, the current case fatality rate was taken from the website of the Oxford

COVID-19 Evidence Service by the Centre for Evidence-Based Medicine (OCEM) at the University of Oxford[1], to provide participants with an accurate figure. During the time of data collection for our first study there was substantial uncertainty around the accuracy of reported COVID-19 case fatality rate estimates due to a variety of reasons influencing the quality of the underlying evidence, as outlined for example on the OCEM website. Sources of uncertainty included, for instance, difficulty determining the exact number of COVID-19 infections due to limitations in testing, or issues with the attribution of deaths to coronavirus (association-causation problem) and how exactly the numbers should be calculated. The available information on the types of issues affecting the quality of the evidence underlying case fatality rate estimates was distilled into our experimental manipulations of uncertain quality of evidence, namely, lack of data and disagreement between experts. For experimental purposes, in Studies 2 and 3 we vary the quantitative aspect of the quality-of-evidence information (i.e., the label for the quality level – high/low) — each compatible with the situation regarding the quality of the evidence base in the recent past. The qualitative aspects (i.e., the source – lack/availability of data and expert disagreement/agreement) remain consistent across studies.

At the time of data collection for Study 1, the case fatality rate in the UK was fairly uncertain, but it was more certain at the time of data collection for Studies 2 and 3. All participants received a debrief note in which this was clarified. They also received a note explaining that reported case fatality rates depend on aspects such as amount of testing that is conducted and were provided with the link to the OCEM website for further reading.

Apart from information from the OCEM website, the two experimental manipulations of sources of uncertainty in evidence quality used in this research, were also based on assessment of factors that contribute to ratings of quality of evidence by groups such as the Intergovernmental Panel on Climate Change (IPCC, 2014), who outline that the confidence in the validity of a finding is based, amongst other factors, on the amount of evidence and the degree of agreement. Expert disagreement and insufficient scientific evidence have also been highlighted as sources of uncertainty in health care (e.g., Han et al., 2011). Finally, conflict and consensus (akin to 'expert (dis) agreement') and ambiguity (represented by our label 'lack of data') have previously been investigated as factors of interest by psychologists studying risk perception and decision making (Baillon et al., 2012; Benjamin & Budescu, 2018; Cabantous, 2007; Cabantous et al., 2011; Koehler & Pennycook, 2019; Roozenbeek & van der Linden, 2019; Smithson, 1999, 2015; van der Linden et al., 2015).

Studies 1 and 2 informed the design of Study 3 which constitutes our main, confirmatory round of data collection. All three studies were pre-registered online on the Open Science

---

[1] https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/. Note that the website has been updated continuously. While this allowed us to provide participants with current figures at the time of data collection for our three studies, it also means that the information we used at the various data collection stages is not available any more. It is furthermore unclear for how long the website will remain accessible. We have included screenshots of the OCEM page at the time of our three studies as well as URLs to the historic pages via an internet archive in the supplementary materials.

Framework (Study 1: https://osf.io/a3r6c, Study 2: https://osf.io/2f8dj, Study 3: https://osf.io/p9kbr). Data for all three studies were collected on the Prolific academic (prolific.co) and Respondi platforms (respondi.com), with half the sample for each study being collected on each platform.[2] All three samples constituted of UK residents and were national quota samples matched on age and gender to characteristics of the UK population.

2 Study 1

2.1 Methods

We start by assessing the effects of low quality-of-evidence cues in a three-group between-subject design. In total, 2,099 responses were collected between April 9-11th 2020. As per the pre-registration, participants who failed an attention check[3] were excluded from the sample, resulting in a total sample of $n = 1,942$ for analysis purposes (47.68% male, 52.01% female, M_{age}=45.61; see supplementary materials for detailed demographic composition). The study was powered (a priori) at 95%, alpha level 0.01 for small effects (f=0.1) (see pre-registration for more details).[4]

Experimental conditions.   Participants were randomly allocated to one of three conditions, a control group and two low quality-of-evidence conditions.[5] All participants were presented with information on the COVID-19 case fatality rate which read: *"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."* Participants in the control group received no further information. Participants in the two low quality conditions in addition either read *"The quality of the evidence underlying the reported case fatality rate is uncertain, because there is disagreement between experts"* (Low-disagree) or *"The quality of the evidence*

---

[2]We conducted checks to investigate a potential influence of the two different sampling platforms on our experimental effects across all three studies. Experimental effects stayed significant controlling for platform, and no interactions were observed. See supplementary materials for detailed analysis results.

[3]The attention check was a multiple choice question asking participants to select the appropriate case fatality rate they had been presented with (see supplementary materials for exact wording). The same attention check (with case fatality rates specific to each study) was used for all three studies.

[4]For power calculations an adjusted alpha level of 0.01 was used to ensure adequate power for the circumstance of low inter-correlations between our items and hence the need to treat them all as separate dependent measures. Our dependent measures ended up being sufficiently correlated to pool them into index measures. Thus the alpha level for analysis purposes remained at 0.05 and the slightly lowered sample size due to exclusions did therefore not adversely affect the study's power.

[5]As pre-registered, we ran balance checks for all three studies to test whether random assignment successfully balanced background variables (i.e., gender, age, education, and political orientation) across experimental conditions. Where one of the variables was associated to the independent variables, we controlled for it in our models. Imbalances were observed for education in Study 1 and political orientation in Study 2. All observed effects remained significant after controlling for these variables. Detailed analyses are reported in the supplementary materials.

underlying the reported case fatality rate is uncertain, because there is a lack of data"[6] (Low-lack).

Measures.   After being presented with the case fatality rate intervention text, participants completed a variety of measures pertaining to the presented information. They indicated to what extent they thought that the COVID-19 case fatality rate mentioned was certain, on a 7-point scale from 'not certain at all' to 'very certain'. This measure was used as a mediator for an exploratory mediation analysis. They also indicated their level of perceived trustworthiness of the presented information, which was measured via three items which were subsequently merged into an index measure ($\alpha = 0.94$). The three items were: "*To what extent do you think that the COVID-19 case fatality rate mentioned is accurate?*", "*To what extent do you think that the COVID-19 case fatality rate mentioned is reliable?*", "*To what extent do you think the COVID-19 case fatality rate mentioned is trustworthy?*", all measured on a 7-point scale from 'not accurate/reliable/trustworthy at all' to 'very accurate/reliable/trustworthy'. These three items encompass the dimensions of trustworthiness as suggested by O'Neill (2018) which include competence and reliability. Participants subsequently responded to questions regarding intended use of the presented information for decision making. Two items, which were combined into an index as pre-registered ($\alpha = 0.75$)[7], probed personal use in decision making as well as government use in decision making: "*How likely are you to base your own COVID-19 related decisions and behaviours on the mentioned case fatality rate*" (7-point scale ranging from 'not at all likely' to 'very likely') and "*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?*" (7-point scale ranging from 'not at all' to 'very much'). Participants were also asked how much they thought was currently known about COVID-19. This measure however was found to be rather uninformative due to problems with its design. We therefore present all analyses for the knowledge measure, together with a discussion of the measure's design issues, in the

---

[6]Note that while we used the wording 'uncertain', we imply low quality of evidence by providing the lack of data and disagreement explanations. To ensure that participants indeed interpreted this wording to mean low quality of evidence, we explicitly compare the wording 'uncertain' and 'low' in two experimental groups in Study 3. We do not find a significant difference between the two conditions with regards to people's reactions. Therefore, for clarity and ease of reading purposes we refer to the two conditions as 'low' quality-of-evidence groups. Refer to the supplementary materials for full analysis results. Given this empirical evidence that people seem to not distinguish 'uncertain' and 'low' quality of evidence in the context of our study in their interpretation, we assume that the same holds for 'fairly certain' and 'high' quality of evidence wording. Indirect evidence in support of this supposition is provided by our findings in studies 2 (where we used a 'fairly certain' wording for the high quality of evidence conditions) and 3 (where we used a 'high' wording for the high quality of evidence conditions) which are overall consistent as to their effects vis-à-vis the control and ambiguous quality-of-evidence groups.

[7]In our pre-registrations for the three studies, we specified a correlation as the threshold metric to use for aggregation of individual items into indices. During the review process it was suggested that we use Cronbach's alpha throughout for consistency. Reliability of the scales used in this research can be considered good by conventional standards.

supplementary materials, as suggested by the reviewers.[8]

2.2 Results and Discussion

2.2.1 Effect of uncertainty information on trust and decision making

We investigate experimental effects on perceived trustworthiness and intended use in decision making. Sample means and confidence intervals per experimental group are visualized in Figures 1 and 2 (refer to the supplementary materials Table S2 for exact means and standard deviations per experimental group). Consistent with our pre-registration, a one-way analysis of variance (ANOVA) indicated a significant difference between experimental groups for both perceived trustworthiness ($F[2, 1933] = 15.02$, $p < .001$, $\eta_p^2 = 0.015$) and decision making ($F[2, 1933] = 8.26$, $p < .001$, $\eta_p^2 = 0.008$). Pre-registered post-hoc testing using Tukey HSD (Table 1) found that participants who were presented with low quality-of-evidence information as indicated through disagreement between experts (Low-disagree) perceived the information as significantly less trustworthy than a control group which did not receive any quality-of-evidence indication (Control) and indicated lower use of the information in decision making.[9] We find the same pattern of results for the group who was presented with low quality-of-evidence information as indicated through a lack of data (Low-lack). These participants perceived the information as significantly less trustworthy than participants in the control group and they indicated lower use of the information in decision making. No significant difference was observed between the two low quality-of-evidence conditions, either for perceived trustworthiness or decision making.

Results from Study 1 indicate that people react to cues that imply low quality of the underlying evidence. We find that people in the low quality-of-evidence groups perceived the information given to them as significantly less trustworthy and indicated significantly lower use in self- and other-oriented decision making compared to a control group.

TABLE 1: Study 1 pairwise comparisons — effect sizes and significance.

Contrast	Perceived trustworthiness		Use in decision making	
	d	p	d	p
Control vs. Low-disagree	**0.24**	**< 0.001**	**0.20**	**0.001**
Control vs. Low-lack	**0.28**	**< 0.001**	**0.19**	**0.002**
Low-disagree vs. Low-lack	0.05	0.598	0.01	0.963

[8]Note that for both Studies 1 and 2 several other outcome measures were collected. For clarity and comparability with Study 3 — our main confirmatory study — only those measures that were used in Study 3 are presented here for Studies 1 and 2. For a full list of all measures included in the surveys as well as their analyses, refer to the supplementary materials.

[9]Given mild skew in the distribution of the decision making outcome variable, we performed a non-parametric robustness check analysis. Results were in line with the presented parametric findings.
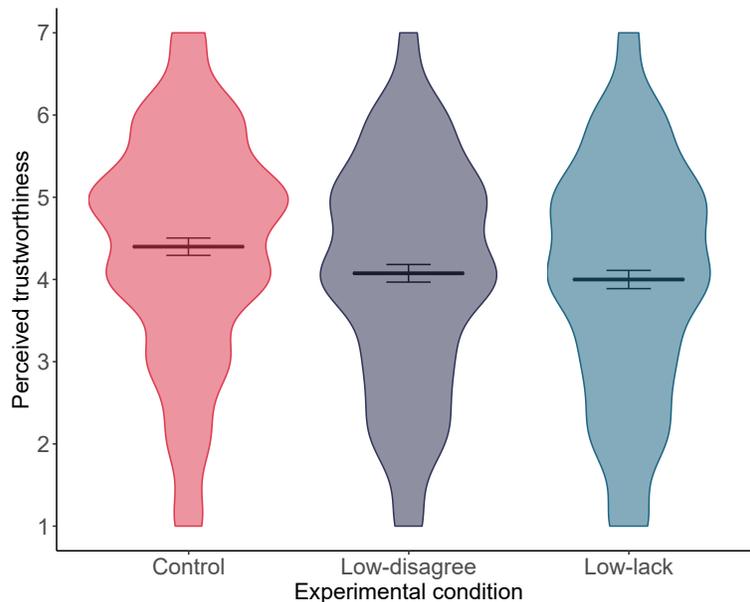
857

FIGURE 1: Experimental effects of low quality of evidence as indicated through disagreement between experts (middle) and lack of data (right) versus control (left) on perceived trustworthiness of the information. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.
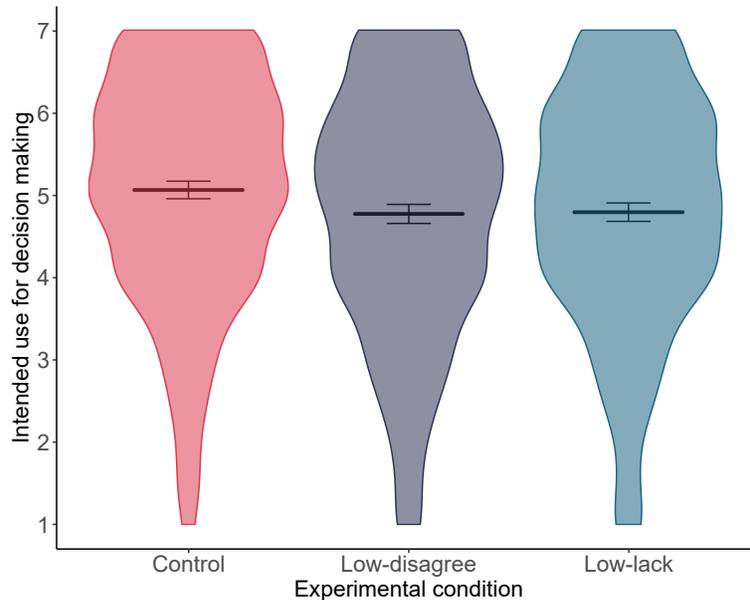

FIGURE 2: Experimental effects of low quality of evidence as indicated through disagreement between experts (middle) and lack of data (right) versus control (left) on intended use of the information for decision making. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.

858

2.2.2 Mediation analysis

We use exploratory structural equation modelling (SEM) to investigate whether perceived uncertainty mediates the effect of quality-of-evidence information on perceived trustworthiness, which in turn influences decision making, including latent variables (Figure 3).[10]

We use this serial approach to estimate the indirect effects using multiple sequential mediators. The experimental conditions act as an exogenous variable. To implement our structural equation models in R, we use the *lavaan* package (Rosseel, 2012), employing maximum likelihood estimation, with bootstrapped confidence intervals for the indirect effects (1,000 samples; using bias corrected adjusted bootstrap percentile method). In order to keep the amount of comparisons manageable, we limit ourselves to testing only those contrasts for which we find a significant direct effect of experimental condition on decision making within the SEM regression framework. Additionally, testing mediation for contrasts without a significant direct effect is considered less convincing. See supplementary

---

[10]We note that we had pre-registered separate mediation analyses of perceived uncertainty on trustworthiness and on decision making (mediations were pre-registered for Studies 2 and 3, and exploratory in Study 1). Looking at our experimental patterns for trustworthiness and decision making descriptively, we observed that results for decision making followed results for trustworthiness, but in a less pronounced way, suggesting that possibly trustworthiness might be acting as a mediator on decision making. This suggested pattern is supported by findings that showed that perceived trustworthiness mediated the effect of quality of evidence information on uptake intentions of health protective behaviours (Schneider et al., 2021). We therefore also investigated trustworthiness as a mediator on decision making in an exploratory fashion, hypothesizing that perceived trustworthiness of the information mediates the experimental effects on use in decision making. During the review process, the reviewers suggested bringing the different theoretical models and hypotheses together, and modelling the relationships more holistically in a structural equation model rather than in separate mediation models. Based on the literature and empirical evidence presented in the introduction, as well as our findings in the present research that uncertainty mediated the effect on trustworthiness and that trustworthiness mediated the effect on decision making, we deemed a serial mediation model in which quality of evidence information influences perceived uncertainty which affects perceived trustworthiness, which in turns influences decision making as most appropriate. This sequence is further supported by the causal ordering of measures in our main confirmatory Study 3, where participants were presented with the quality of evidence manipulation in the various experimental groups, followed by indicating their level of perceived uncertainty of the information, their perception of trustworthiness of the information, and finally responded to the decision making items. We implement this exploratory structural equation model across our three studies. We observe good model fit across all three studies. Nevertheless, given the exploratory nature of the implemented path model, we additionally explore two alternative models, as suggested during the review process: one alternative serial model in which perceived uncertainty and trustworthiness are switched in sequence, as well as a parallel model in which both mediators jointly affect decision making. Results of these additional exploratory models are presented in the supplementary materials, along with comparisons to the main model presented here. Apart from describing descriptive differences in results – i.e., (non-)significance of investigated indirect effect paths – and model fit indices, we evaluate the plausibility of the competing path models on the basis of Akaike (AIC) and Bayesian (BIC) Information Criteria, employing rules of thumb based on Burnham and Anderson (2004) and Burnham, Anderson and Huyvaert (2011) for AIC differences between models, and Raftery (1995, 1999) for BIC differences. The balance of evidence speaks in favour of the main model presented here suggesting that overall, across our three studies and contrasts, the main serial model seems to be a better description of the mechanism at play compared to the alternative models. We remind the reader though of the exploratory nature of the conducted structural equation modelling, and hence refrain from making definitive claims. We instead encourage future research — employing confirmatory methods — for investigating the relationships at play in more detail. See supplementary materials for more details.
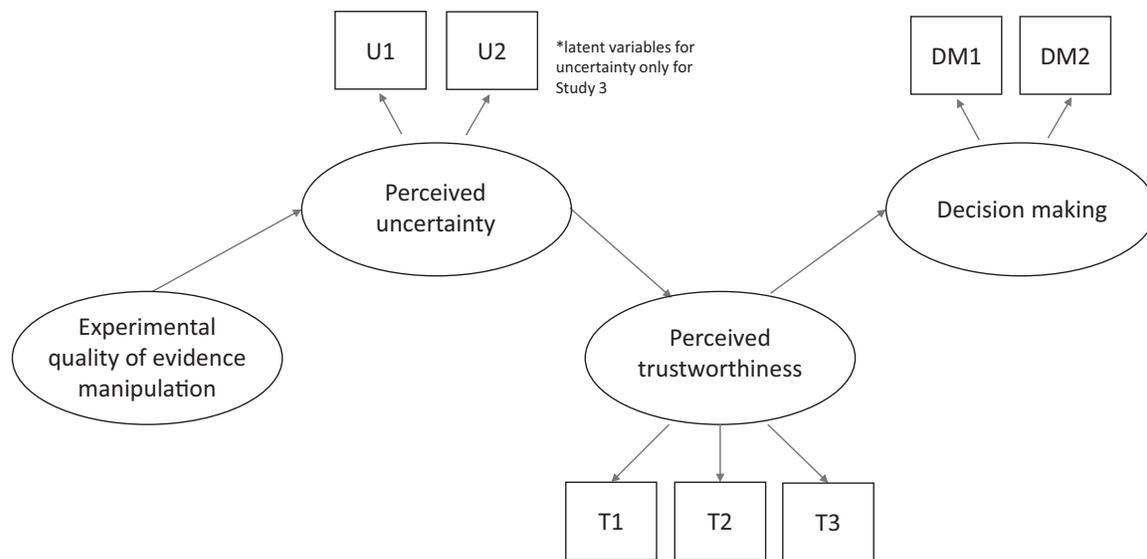
859

FIGURE 3: Structure of proposed serial mediation model. Model consisting of only the indirect effect paths. Effect directions throughout the model are hypothesized as follows: for experimental contrasts that lead to an increase [a decrease] in perceived uncertainty, we hypothesize a decrease [an increase] in perceived trustworthiness, which will in turn decrease [increase] indicated use in decision making. Latent variables are modelled for perceived trustworthiness and decision making for all three studies; and for perceived uncertainty for Study 3.

materials for direct effects output tables. For Study 1 we formally test the following two contrasts: Control vs. Low-disagree and Control vs. Low-lack.

Results for Study 1 (Table 2) indicated good model fit to the data for both examined contrasts according to conventional guidelines; RMSEA and SRMR fit indices are below 0.06 and 0.08 respectively, and CFI indices are above 0.95 which indicates excellent fit (Hu & Bentler, 1999; Lei & Wu, 2007). We note that the χ^2 goodness of fit test was significant for both contrasts. However, as this is a test of whether the hypothesized covariance structure is different from the real covariance structure (test of perfect fit), tiny and trivial variations might lead to significant test results for big sample sizes, despite the model fitting the data reasonably well. As our samples sizes are above 600 per experimental group it is likely that observed significant effects are due to sample size rather than an indicator of concerning levels of differences between observed and model-implied covariance matrices (Bentler, 1990; Hu et al., 1992; Lei & Wu, 2007; McNeish, 2018). Taken together, we thus believe the observed fit indices indicate good fit of the theoretically derived model.

Indirect effects of all tested contrasts emerged as significant (bootstrapped confidence intervals not including zero). Results support the theoretical model laid out, namely that uncertainty mediates the relationship between quality-of-evidence information and perceived trustworthiness, and that trustworthiness, in turn, predicts decision making. Concretely, observed results suggest that being presented with low quality-of-evidence information, either

TABLE 2: Structural equation model indirect effects and model fit for Study 1.

	Indirect effect		Model fit							
Contrast	Estimate	95% CI	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Low-disagree	–0.16	[–0.23, –0.09]	43.21	13	< .001	0.995	0.042	0.017	22649.30	22721.56
Control vs. Low-lack	–0.20	[–0.28, –0.12]	63.61	13	< .001	0.992	0.055	0.017	23006.82	23079.20

due to disagreement between experts (Low-disagree) or lack of data (Low-lack), compared to information that did not include a quality-of-evidence indicator (Control), increased people's perceived uncertainty of the information, which decreased perceived trustworthiness of the information, and in turn decreased use in decision making.

Our path analysis results point towards an increase in perceived uncertainty of the information as a potential cause for the observed effects on trustworthiness and decision making. Given the downward effects of low quality of evidence we wanted to test the extent of possible upward effects for high quality of evidence. We furthermore wanted to investigate how people would react to ambiguous quality-of-evidence cues. We hypothesized that we would find downward effects, given the literature on ambiguity aversion. We ran Study 2 to tackle these research questions.

3 Study 2

3.1 Methods

In Study 2, we therefore assess the effects of high and ambiguous quality of evidence. In total, 2,309 responses were collected between May 7-11th 2020. The final analytic sample post attention check failure exclusion was $n = 2,155$ (48.21% male, 51.6% female, M_{age}=45.47; see supplementary materials for detailed demographic composition). Like Study 1, power calculation was based on a conservative alpha level of 0.01, estimated small effect size (f=0.1), and 95% power, to include a buffer for attrition due to exclusions.

Experimental conditions. Participants were randomly allocated to one of four conditions, a control group, two high quality-of-evidence conditions, and an ambiguous quality-of-evidence condition. All participants were presented with information on the COVID-19 case fatality rate, with the same wording as in Study 1 (the actual rate information was updated to reflect the correct case fatality rate at the time of data collection). As in Study 1, participants in the control group received no further information. Participants in the two high quality cue conditions in addition either read *"The quality of the evidence underlying the*

861

reported case fatality rate is fairly certain, because there is a high level of expert agreement" (High-agree) or *"The quality of the evidence underlying the reported case fatality rate is fairly certain, because there is a large amount of data available"*[11] (High-data). Participants in the ambiguous quality-of-evidence condition further read: *"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"* (Ambiguous).

Measures. Using the same measures as in Study 1 participants indicated their perceived uncertainty of the presented information – for use as a mediator in a pre-registered mediation analysis, their level of perceived trustworthiness ($\alpha = 0.95$) and their intended use of the presented information for decision making ($\alpha = 0.79$).

3.2 Results and Discussion

3.2.1 Effect of uncertainty information on trust and decision making

As for Study 1, we investigate differences in perceived trustworthiness and intended use of the provided information for decision making across experimental groups. Sample means and confidence intervals per experimental group are visualized in Figures 4 and 5 (refer to the supplementary materials Table S4 for exact means and standard deviations per experimental group). Our pre-registered one-way analysis of variance revealed a significant main effect of experimental condition on perceived trustworthiness ($F[3, 2149] = 36.23$, $p < .001$, $\eta_p^2 = 0.048$). Specifically, post-hoc comparisons using Tukey HSD as pre-registered (Table 3) revealed that participants who were presented with ambiguous quality-of-evidence information (Ambiguous) perceived the information as significantly less trustworthy than a control group (Control), the high quality-of-evidence group indicated by expert agreement (High-agree), and the high quality-of-evidence group indicated by the availability of ample data (High-data). Additionally, results indicated that people in the high quality-of-evidence group as indicated through expert agreement (High-agree) perceived the information as significantly more trustworthy compared to control group participants (Control). However, people in the high quality-of-evidence group as indicated through availability of data (High-data) did not perceive it as significantly more trustworthy compared to people in the control group (Control). No statistically significant difference was observed between the two high quality-of-evidence groups.

Results for use of the information in decision making revealed a similar pattern ($F[3, 2149] = 16.86$, $p < .001$, $\eta_p^2 = 0.023$).[12] Participants in the ambiguous quality-of-evidence

---

[11]Ethically we could not use the phrase 'certain' as it would have been misleading to participants given the actual current state of knowledge of the case fatality rate. We therefore use the ethically appropriate phrasing of 'fairly certain'.

[12]Given mild skew in the distribution of the decision making outcome variable, we performed a non-parametric robustness check analysis. Results were in line with the presented parametric findings.
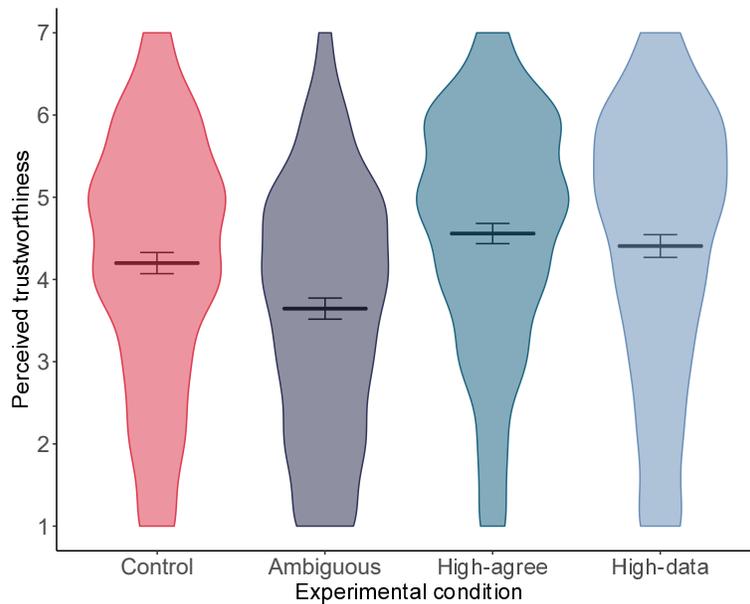
FIGURE 4: Experimental effects of high (agreement and availability of data) and ambiguous quality of evidence versus control on perceived trustworthiness of the information. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.
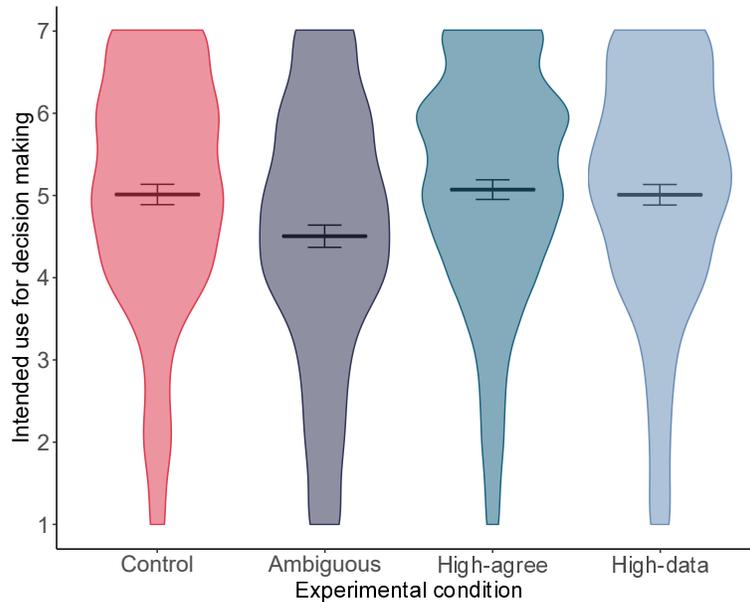

FIGURE 5: Experimental effects of high (agreement and availability of data) and ambiguous quality of evidence versus control on intended use of the information for decision making. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.

TABLE 3: Study 2 pairwise comparisons — effect sizes and significance.

Contrast	Perceived trustworthiness		Use in decision making	
	d	p	d	p
Control vs. Ambiguous	**0.36**	**< 0.001**	**0.33**	**< 0.001**
Control vs. High-agree	**0.24**	**0.001**	0.04	0.914
Control vs. High-data	0.13	0.115	< 0.01	1
Ambiguous vs. High-agree	**0.62**	**< 0.001**	**0.38**	**< 0.001**
Ambiguous vs. High-data	**0.48**	**< 0.001**	**0.33**	**< 0.001**
High-agree vs. High-data	0.10	0.363	0.04	0.899

group (Ambiguous) also indicated significantly lower use of the information in decision making compared to the control group (Control), the high quality-of-evidence group indicated through expert agreement (High-agree), and the high quality-of-evidence group indicated through ample availability of data (High-data). As opposed to trustworthiness, there was no significant difference between the high quality-of-evidence group indicated through expert agreement (High-agree) and the control group (Control). In line with trustworthiness findings, there was no significant difference between the high quality-of-evidence group as indicated by data availability (High-data) and the control group (Control), nor between the two high quality-of-evidence groups.

Findings from Study 2 confirmed our hypotheses regarding the ambiguous quality-of-evidence condition: We find a strong downward effect, such that people perceived the information to be less trustworthy and indicated lower use in decision making when they were presented with ambiguous quality cues compared to a control group. For high quality of evidence our results were somewhat mixed. Findings for decision making were in line with our expectations, such that we do not find a significant difference between the two high quality groups and the control group. For perceived trustworthiness we did find a small effect of high quality of evidence as indicated through expert agreement, with people perceiving the information significantly more trustworthy compared to control group participants. However, no significant difference between the high quality group as indicated through availability of data and the control group emerged.

3.2.2   Mediation analysis

As laid out in Study 1, we test whether quality-of-evidence information influences perceived uncertainty, to influence trustworthiness, and in turn decision making, via a structural equation path model. We investigate the following three contrasts for which we observed significant direct effects of experimental condition on decision making: Control vs. Ambiguous, Ambiguous vs. High-agree, and Ambiguous vs. High-data. As for Study 1, model

fit indices overall indicated good fit of our model to the data (Table 4). See the Results section of Study 1 for details on how we evaluated our models.

TABLE 4: Structural equation model indirect effects and model fit for Study 2.

Contrast	Indirect effect		Model fit							
	Estimate	95% CI	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Ambiguous	–0.32	[–0.43, –0.23]	26.58	13	0.014	0.998	0.031	0.020	18998.80	19068.51
Ambiguous vs. High-agree	0.56	[0.45, 0.67]	13.96	13	0.376	1	0.008	0.009	18845.87	18915.46
Ambiguous vs. High-data	0.46	[0.35, 0.58]	26.70	13	0.014	0.998	0.031	0.012	19110.00	19179.71

As in Study 1, indirect effects of all tested contrasts emerged as significant (bootstrapped confidence intervals not including zero), thus supporting our theoretical mediation model. We find that ambiguous quality-of-evidence information (Ambiguous), compared to no quality-of-evidence information (Control), led to an increase in perceived uncertainty, which lowered trustworthiness and in turn use in decision making. Providing people with high quality-of-evidence information, both indicated via agreement between experts (High-agree) and ample availability of data (High-data), compared to ambiguous quality-of-evidence information (Ambiguous), led to a decrease in perceived uncertainty, an increase in perceived trustworthiness of the information and, apparently in turn, an increase in use in decision making.

Path analysis findings replicate insights from Study 1, showing evidence that perceived uncertainty acts as a mediator of the effect of the experimental conditions on perceived trustworthiness, with trustworthiness in turn acting as a mediator of the effect on decision making. Because of the somewhat inconclusive findings around the effect of high quality of evidence and because Study 2 did not allow us to directly compare effects for high and low quality of evidence, we decided to include these conditions in a final confirmatory study. We were furthermore interested in pitting the effects of ambiguous quality of evidence against low quality of evidence. Results from our low quality-of-evidence conditions in Study 1 and from our ambiguous quality-of-evidence condition in Study 2 had revealed strong effects on people's perceptions of trustworthiness and use of the information in decision making; however, given that the low and ambiguous quality-of-evidence conditions were run in two separate studies, we could not compare them directly to each other. Hence, we included the ambiguous quality-of-evidence condition in the final confirmatory study as well.

4   Study 3

4.1   Methods

Our final confirmatory study assesses the effects of low, high, and ambiguous quality of evidence together.[13] In total, 2,651 participants were sampled between July 6–9th 2020. The final analytic sample after exclusions was $n = 2,392$ (47.37% male, 52.38% female, M_{age}=45.16; see supplementary materials for detailed demographic composition). The sample that we had powered for provided more than 95% power at alpha level 0.05 for small effects (f=0.1) and included a buffer to account for attrition, hence our final analytic sample is within the bounds of the power analysis.

Experimental conditions.   Participants were randomly allocated to one of eight conditions, a control group that did not provide any quality-of-evidence indicator, three low quality-of-evidence conditions, two high quality-of-evidence conditions, an ambiguous quality-of-evidence condition, and an additional low quality-of-evidence group serving as a robustness check, described below.[14] As in Studies 1 and 2, all participants received the information on the COVID-19 case fatality rate updated to the time of data collection. Participants in the three low quality conditions in addition read one of the following pieces of text: (1) *"The quality of the evidence underlying the reported case fatality rate is low, because there is disagreement between experts"* (Low-disagree) — which was the same condition as used in Study 1 for replication purposes, except for a slight wording change.[15] The additional low quality-of-evidence robustness check group repeated the exact wording from Study 1 to ensure that there was no significant difference between the two wording versions in terms of their effects on the outcome variables. Analyses confirmed that the two wording conditions were not significantly different. Details on the analysis are presented in the supplementary materials. (2) Another low quality-of-evidence experimental condition replicated the lack of data condition from Study 1 (*"The quality of the evidence underlying the reported case fatality rate is low, because there is a lack of data"*) (Low-data), and (3) the final low quality-of-evidence condition did not give any explanatory details but merely presented the quality-of-evidence level 'label', i.e., that it was 'low' (*"The quality of the evidence underlying the reported case fatality rate is low"*) (Low). The two high quality conditions followed the outlined rationale of the low quality-of-evidence conditions, by

---

[13]Evidence on the case fatality rate at the point of data collection was so variable that all three quality-of-evidence cues (i.e., high, low, ambiguous) could be justified without misleading participants.

[14]In Study 1 we had used the phrasing 'uncertain quality of evidence' (together with a reason for the low quality) instead of explicitly calling it 'low quality of evidence'. In Study 3 we use the more explicit wording for our main experimental groups, but also run an additional experimental group replicating the exact wording from Study 1 to test for potential differences as a robustness check.

[15]In Study 1 the quality-of-evidence level indicator for the disagreement group had used the wording 'uncertain', i.e., 'The quality of the evidence underlying the reported case fatality rate is uncertain, because there is disagreement between experts'. In Study 3 we changed 'uncertain' to 'low' to be as explicit as possible about the low quality-of-evidence level.

866

replicating the condition from Study 2 in one group (*"The quality of the evidence underlying the reported case fatality rate is high, because there is a high level of expert agreement"*) (High-agree) and using a 'no explanation' wording indicating merely that the quality level was 'high' (*"The quality of the evidence underlying the reported case fatality rate is high."*) (High). Finally, the ambiguous quality-of-evidence condition (Ambiguous) was an exact replication of the one used in Study 2, using the exact same text.

Measures. Outcome measures were identical to those presented for Studies 1 and 2 with only minor changes. To improve the reliability and statistical accuracy of the mediation analysis, the mediator (perceived uncertainty of the presented information) was measured via two items, which were presented on a separate page from the other subsequent measures. The two items were combined into an index ($\alpha = 0.94$): *"To what extent do you **think** that the COVID-19 case fatality rate mentioned is certain or uncertain?"* and *"To what extent do you **feel** that the COVID-19 case fatality rate mentioned is certain or uncertain?"*, both measured on 7-point scales, ranging from 'very certain' to 'very uncertain'. Subsequently, participants responded to the three trustworthiness items which were combined into an index as before ($\alpha = 0.95$). Use in personal decision making was probed using the same items as in Studies 1 and 2. Use in government decision making had a slightly altered wording[16]: *"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?"* (7-point scale ranging from 'not at all' to 'very much'). Both items were combined into an index as before ($\alpha = 0.82$).

4.2 Results and Discussion

4.2.1 Effect of uncertainty information on trust and decision making

Our last confirmatory study replicated and extended findings from Studies 1 and 2. We find differences in perceived trustworthiness and intended use of the provided information for decision making between the experimental groups. Sample means and confidence intervals per experimental group are visualized in Figures 6 and 7 (refer to the supplementary materials table S6 for exact means and standard deviations per experimental group). The pre-registered analysis of variance revealed a significant main effect of experimental condition on perceived trustworthiness ($F[6, 2084] = 12.48$, $p < .001$, $\eta_p^2 = 0.035$). As outlined in the pre-registration, Study 3 allowed us to not only compare low and high quality-of-evidence groups to the control group that did not give quality-of-evidence information, but to additionally compare them to the ambiguous quality-of-evidence group as well as to each

---

[16]Here the words 'and recommendations' were added compared to the measure used in Studies 1 and 2. This slight wording change was implemented to make it clearer to the reader that decisions which the government takes often result in recommendations which can affect people's lives; and to offer a counterpart to the 'and behaviours' section in the personal decision making measure.

867

other. As indicated in the pre-registration, we use Tukey HSD for our post-hoc comparisons (Table 5).

FIGURE 6: 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 information. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.
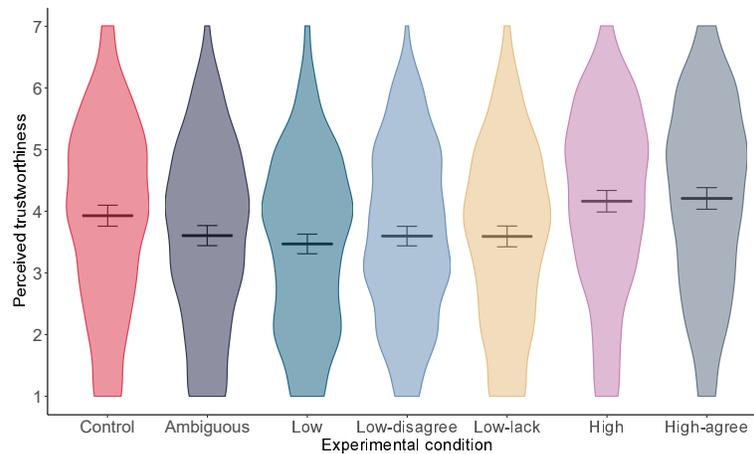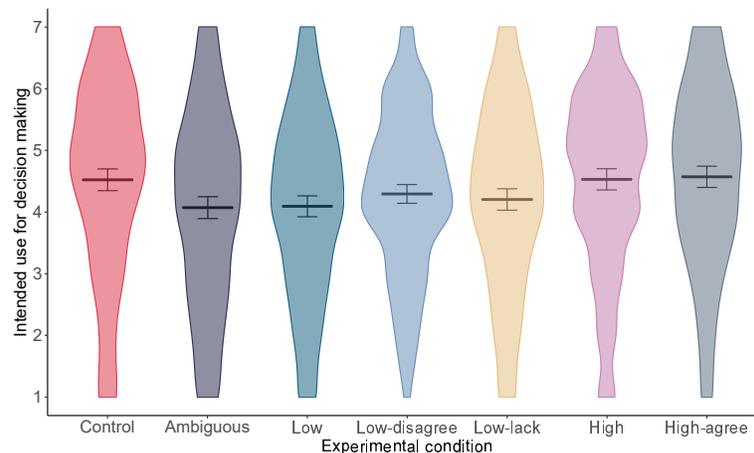

FIGURE 7: Experimental effects of high (agreement and no reason), low (disagreement, lack of data, and no reason), and ambiguous quality of evidence versus control on intended use of the information for decision making. Horizontal lines and error bars denote sample means and 95% confidence intervals respectively. Violin plots visualize data distributions.

Looking at perceived trustworthiness of the information, we find that trustworthiness was lower for participants in the low quality-of-evidence groups (Low, Low-disagree, Low-lack) compared to participants in the control group (Control), in line with Study 1 results. This observed descriptive pattern was statistically significant for the low quality-of-evidence group that did not provide an explanation for the quality label (Low) compared to the

TABLE 5: Study 3 pairwise comparisons — effect sizes and significance.

Contrast	Perceived trustworthiness		Use in decision making	
	d	p	d	p
Control vs. Ambiguous	0.22	0.090	**0.29**	**0.004**
Control vs. Low	**0.31**	**0.002**	**0.28**	**0.007**
Control vs. Low-disagree	0.23	0.083	0.16	0.497
Control vs. Low-lack	0.23	0.076	0.21	0.124
Control vs. High	0.16	0.452	0	1
Control vs. High-agree	0.18	0.225	0.03	1
Ambiguous vs. Low	0.09	0.915	0.01	1
Ambiguous vs. Low-disagree	0.01	1	0.15	0.530
Ambiguous vs. Low-lack	0.01	1	0.09	0.933
Ambiguous vs. High	**0.38**	**< 0.001**	**0.30**	**0.004**
Ambiguous vs. High-agree	**0.40**	**< 0.001**	**0.32**	**0.001**
Low vs. Low-disagree	0.09	0.938	0.14	0.657
Low vs. Low-lack	0.08	0.951	0.07	0.972
Low vs. High	**0.48**	**< 0.001**	**0.29**	**0.008**
Low vs. High-agree	**0.50**	**< 0.001**	**0.32**	**0.002**
Low-disagree vs. Low-lack	0	1	0.06	0.991
Low-disagree vs. High	**0.39**	**< 0.001**	0.17	0.480
Low-disagree vs. High-agree	**0.42**	**< 0.001**	0.20	0.268
Low-lack vs. High	**0.39**	**< 0.001**	0.22	0.120
Low-lack vs. High-agree	**0.41**	**< 0.001**	**0.24**	**0.046**
High vs. High-agree	0.03	1	0.03	1

control group (Control). The two low quality-of-evidence conditions that provided an explanation for the quality label, i.e., disagreement between experts (Low-disagree) and lack of data (Low-lack), were not significantly different from the control group (Control), despite substantial descriptive differences. Expanding on Studies 1 and 2, no significant difference was observed between the ambiguous quality-of-evidence group (Ambiguous) and all three low quality-of-evidence conditions (Low, Low-disagree, Low-lack). In line with results of Study 2, participants in both high quality-of-evidence groups, i.e., high quality of evidence without an explanation (High) and high quality of evidence as indicated by agreement between experts (High-agree), perceived the information as significantly more trustworthy compared to participants in the ambiguous quality-of-evidence group (Ambiguous). Furthermore, the two high quality-of-evidence groups (High, High-agree)

were not significantly different from the control group (Control) in line with Study 2 (with the exception of the significant difference between the control group (Control) and expert agreement group (High-agree) for perceived trustworthiness in Study 2). Building on findings from Studies 1 and 2, a direct comparison between high and low quality-of-evidence conditions revealed a significant difference between both high quality-of-evidence groups and all three low quality-of-evidence groups. In line with the pattern observed in Study 2 we find that participants in the ambiguous quality-of-evidence group (Ambiguous) indicated lower perceived trustworthiness than participants in the control group (Control). However, this difference was non-significant in our post-hoc inferential analysis. Finally, the three low quality-of-evidence groups (Low, Low-disagree, Low-lack) did not significantly differ from each other; nor did the two high quality-of-evidence groups (High, High-agree).

Results for use of the information in decision making likewise revealed a significant main effect of experimental condition ($F[6, 2084] = 6.15$, $p < .001$, $\eta_p^2 = 0.017$), albeit less pronounced than for perceived trustworthiness, in line with findings from Studies 1 and 2. As in Study 1 conditions which presented participants with low quality-of-evidence information (Low, Low-disagree, Low-lack) resulted in lower indicated use of the information in decision making compared to the control group that did not mention quality of evidence at all (Control). However, only the difference between the low quality-of-evidence group that did not provide an explanation (Low) and the control group (Control) was statistically significant, not those between the two low quality-of-evidence groups that provided an explanation (Low-disagree, Low-lack) and the control group (Control). As for perceived trustworthiness, no significant difference was observed between the ambiguous quality-of-evidence group (Ambiguous) and all three low quality-of-evidence groups (Low, Low-disagree, Low-lack). Consistent with results for perceived trustworthiness in Study 3 and with decision making findings in Study 2, participants in both high quality-of-evidence conditions (High, High-agree) indicated significantly higher use of the information in decision making compared to participants in the ambiguous quality-of-evidence condition (Ambiguous). The two high quality-of-evidence conditions (High, High-agree) were not significantly different from the control condition (Control), in line with perceived trustworthiness findings in Study 3 and results for decision making in Study 2. Although both high quality-of-evidence conditions (High, High-agree) were descriptively higher on our measure of use in decision making than all three low quality-of-evidence conditions (Low, Low-disagree, Low-lack), unlike for trustworthiness, this difference was only statistically significant for some of the contrasts. In line with Study 2 findings for decision making, participants in the ambiguous quality-of-evidence condition (Ambiguous) indicated descriptively and inferentially lower use of the information for decision making compared to participants in the control condition (Control). Finally, the three low quality-of-evidence groups (Low, Low-disagree, Low-lack) did not significantly differ from each other; nor did the two high quality-of-evidence groups (High, High-agree).

In addition to replicating findings from Studies 1 and 2 with regards to the experimental

870

effects of high quality-of-evidence, low quality-of-evidence, and ambiguous quality-of-evidence information against a control group that did not provide any quality-of-evidence indicator, Study 3 also allowed us to directly compare effects of providing high versus low quality-of-evidence information with each other, as well as low and ambiguous quality of evidence. We find that people react more strongly to low quality-of-evidence cues compared to high quality-of-evidence cues when comparing to a control group that did not receive any quality-of-evidence cue. We find a significant decline in perceived trustworthiness and use in decision making for low quality-of-evidence information that did not provide an explanation compared to the control group that did not give quality-of-evidence information. No significant differences between the high quality-of-evidence groups and the control group emerged, neither for perceived trustworthiness nor for decision making. Furthermore, insights from Study 3 reveal that the decrease in trustworthiness and use in decision making for low quality-of-evidence cues (compared to control) is about as pronounced as the decrease from ambiguity regarding the quality level (compared to control); we do not find a significant difference between the three low quality-of-evidence groups and the ambiguous quality-of-evidence group for any of our outcome measures. Comparing the three low quality-of-evidence groups against each other, and comparing the two high quality-of-evidence groups to each other, we do not find significant differences, suggesting that the addition or omission of a reason for the quality level, as well as the exact nature of the reason, did not seem to play a role in this context.

4.2.2 Mediation analysis

We test our structural equation path model for the following eleven contrasts (out of the twenty-one total contrasts) for which we observed significant direct effects of experimental condition on decision making: (1) Control vs. Ambiguous, (2) Control vs. Low, (3) Control vs. Low-lack, (4) Ambiguous vs. High, (5) Ambiguous vs. High-agree, (6) Low vs. High, (7) Low vs. High-agree, (8) High vs. Low-disagree, (9) High vs. Low-lack, (10) Low-disagree vs. High-agree, and (11) High-agree vs. Low-lack. As for Studies 1 and 2, model fit indices of our exploratory structural equation models were overall good (Table 6). For several contrasts the χ^2 goodness of fit tests were significant (refer to the Results section of Study 1 for a discussion of the significance of a significant goodness of fit test). While sample sizes are lower in Study 3 compared to Studies 1 and 2, they are at around 300 per experimental group, thus large enough to allow to draw conclusions and gather insights from our models by conventional standards.[17] Indirect effects of all tested contrasts emerged as significant (bootstrapped confidence intervals not including zero), providing further support for our theoretical mediation model. Low and ambiguous quality-of-evidence information led to an increase in perceived uncertainty, and consequently a decrease in trustworthiness and decision making compared to control. High quality-of-

---

[17]Hu and Bentler (1999), evaluating robustness of model fit in the SEM context, indicate less reliability at sample sizes of *N*=250 and below. McNeish (2018) defines small samples as *N* < 200.

evidence information led to a decrease in perceived uncertainty, which translated into an increase in perceived trustworthiness and decision making compared to low and ambiguous quality-of-evidence information.

TABLE 6: Structural equation model indirect effects and model fit for Study 3.

	Indirect effect		Model fit							
Contrast	Estimate	95% CI	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
Control vs. Ambiguous	−0.36	[−0.50, −0.23]	39.18	18	0.003	0.995	0.044	0.025	12177.26	12252.43
Control vs. Low	−0.35	[−0.50, −0.21]	22.83	18	0.197	0.999	0.021	0.017	11944.12	12019.09
Control vs. Low-lack	−0.34	[−0.47, −0.22]	32.96	18	0.017	0.996	0.037	0.026	11800.02	11874.68
Ambiguous vs. High	0.55	[0.41, 0.71]	24.34	18	0.144	0.998	0.024	0.024	11889.21	11963.81
Ambiguous vs. High-agree	0.57	[0.43, 0.72]	49.08	18	< .001	0.993	0.053	0.022	11973.94	12048.80
Low vs. High	0.57	[0.44, 0.73]	22.07	18	0.229	0.999	0.020	0.017	11660.93	11735.33
Low vs. High-agree	0.57	[0.43, 0.72]	29.70	18	0.040	0.997	0.033	0.012	11732.32	11806.99
High vs. Low-disagree	−0.51	[−0.65, −0.40]	36.71	18	0.006	0.995	0.042	0.036	11442.74	11516.91
High vs. Low-lack	−0.52	[−0.65, −0.40]	30.10	18	0.037	0.997	0.034	0.030	11564.23	11638.32
Low-disagree vs. High-agree	0.52	[0.40, 0.66]	41.62	18	0.001	0.994	0.047	0.032	11510.92	11585.36
High-agree vs. Low-lack	−0.53	[−0.67, −0.41]	22.22	18	0.223	0.999	0.020	0.023	11600.24	11674.59

Our mediation analysis confirms findings from Studies 1 and 2, highlighting perceived uncertainty as a mediator in the effect of quality-of-evidence information on perceived trustworthiness, and, in turn, trustworthiness as a mediator in the effect on decision making.

5   General Discussion

5.1   Summary of main findings

Across three studies we provide insights into the effects of communicating indirect scientific uncertainty on perceived trustworthiness and decision making in a public health context. Comparing the communication of high and low quality of evidence, high quality-of-evidence cues lead people to perceive the information presented as more certain, more trustworthy, and use it more in decision making, while cues of low quality of evidence lead to lower perceptions and judgments of these. When comparing effects to a control group that was not provided with any quality-of-evidence information, we find an asymmetry in people's responses: effects for cues of low quality of evidence (distance from the control group in the negative direction) are stronger compared to effects for cues of high quality of evidence (distance from the control group in the positive direction). Furthermore, ambiguous quality-of-evidence information leads to a decrease in perceived trustworthiness and use in decision making, about as pronounced as effects for low quality-of-evidence information.

5.2   Interpretation of main findings

Our findings regarding the role of high versus low quality-of-evidence information are in line with recent work on the public's reactions to cues of evidence quality related to non-pharmaceutical public health interventions (Schneider et al., 2021). One potential explanation of these results is that people value the potential gain from cues of high quality evidence less than that from the presence of a low quality cue (versus no quality indication). In other words, presenting people with cues of low quality of evidence may lead to a stronger decline in perceived trustworthiness and indicated use of the information in decision making than the increase in trustworthiness and intentions when presented with cues of high quality of evidence, taking the control group (i.e., not being given any quality-of-evidence indication) as a reference point. Relatedly, work on intuitive judgment looking at the relationship between strength of an effect and weight of the evidence, has found that people pay insufficient regard to high 'credence' of the evidence when the size of an effect is not particularly strong (Griffin & Tversky, 1992).

An alternative explanation is that participants in the control group, with no cues, interpret the information provided as based on high quality evidence: they have a prior that scientific-sounding evidence presented to them is likely of high quality. This would be in line with classic work on how mental systems believe, which argues that humans rapidly and automatically believe in information they understand, similar to how they believe in physical objects they see, and consequently behave as if that automatic belief is true (Gilbert, 1991; Gilbert et al., 1993). So, in the absence of any cue as to the quality of the evidence that the case fatality rate information is based on, people automatically believe the information because they comprehended it easily, and act accordingly.

873

It is important to note that the experimental information given to participants did not mention any source, in order to avoid potential confounds stemming from attitudinal priors towards a particular provider of information. In contexts where information is tied to a source people may be more sceptical of the information depending on their priors, even without any quality-of-evidence indicator. Individual differences, such as pre-dispositions for science scepticism may additionally play a moderating role (Brzezinski et al., 2021). Furthermore, how exactly people interpret missing quality-of-evidence information may be context-dependent. For instance, in a context where it is the norm that quality-of-evidence information is provided, i.e., where the concept of quality of evidence is very internally salient to people and where there is an expectation to receive it, missing quality-of-evidence indication may be interpreted as a 'red flag' and cause people to behave more as they did in the ambiguous quality-of-evidence condition in our study. Future studies should investigate the relationship between people's reactions to high quality-of-evidence indicators versus lack of any mention of the quality of evidence, potential automatic appraisal, the ease of comprehension of the information, as well as context-dependency of effects and contextual aspects such as the role of source and potential moderating influences in more detail. These would allow exploration of whether — and if so to what extent and in what ways — effects depend on a particular context or reflect attributes of quality-of-evidence perception more broadly.

Findings from our analysis of the ambiguous quality-of-evidence conditions confirm our predictions stemming from the literature on ambiguity aversion (Keren & Gerritsen, 1999; Machina & Siniscalchi, 2014), such that people react in a downward direction to ambiguous quality information. Levels of trustworthiness and use in decision making are about equivalent between the low quality and ambiguous quality conditions. Technically, in the ambiguous quality condition the quality-of-evidence level could be higher than in the low quality conditions. Nevertheless, ambiguous quality-of-evidence information leads to as steep a decline in perceived trustworthiness and indicated use in decision making compared to control as the 'surely' low quality conditions.

Following from the work on mental systems (Gilbert, 1991; Gilbert et al., 1993), it may be that while people do not routinely question the reliability of information they are presented with (note that in the control group the quality of the evidence was not mentioned at all, therefore people might not think of it as a factor to consider when evaluating the presented information), the ambiguous quality-of-evidence information highlights and reminds people that information can theoretically be unreliable. From this perspective, if people presume that information is provided only when it is reliable, an unexpected cue that highlights the ambivalent quality of the evidence is informative to people because it might signal that the quality may be lower than expected.

5.3   Considerations for practice

Our findings suggest that the omission of any cues of quality of evidence on public health-related claims and statistics could lead people to trust the information more than might be warranted if the evidence is actually of low quality. Likewise, if people are presented with ambiguous quality-of-evidence cues they might discount the information.

How communicators choose to consider these implications and approach the communication of quality of evidence, may depend on the goal of the communications, namely, whether simply to inform or to persuade (change behaviour). In order to persuade, communicators might feel it is still ethical to omit the disclosure of low or ambiguous quality of evidence. For instance, in public health contexts where communicators might feel that the goal of persuading as many people as possible to adopt certain protective health measures justifies the omission of a classification of low or ambiguous quality of evidence. This might particularly be the case if the evidence quality is low because of a lack of data rather than if there is expert disagreement or conflicting data, and when the associated potential harms of adopting such measures are low. If the goal is purely to inform, however, communicators should aim to disclose uncertainties and honestly acknowledge limitations in scientific knowledge. This has been argued to be vital for informed decision making and the retention of public trust in the long run (Han, 2013; Kreps & Kriner, 2020; Veit et al., 2021). Further research on exactly what the public interpret terms such as 'low' and 'high' quality of evidence to mean will help clarify whether the cues are in fact honestly and correctly communicating the quality of the current state of the knowledge.

5.4   Potential underlying mechanism

Our work also provides initial insights into a potential underlying mechanism of the observed effects of quality-of-evidence communication. Replicated across all three of our studies we find that quality-of-evidence indicators shape how certain or uncertain people perceive the information to be, which influences their level of perceived trustworthiness, and in turn affects use of the information in decision making. Due to the exploratory nature of the full serial path model, we also investigated alternative viable models (presented and discussed in the supplementary materials; see also footnote 10). The balance of evidence — based on model fit evaluation using conventional rules of thumb and taking (non-)significance of indirect effect paths into account — overall speaks in favour of the serial model presented here, in which quality-of-evidence information affects perceived uncertainty, which affects trustworthiness, and in turn decision making. However, given the well-known limitations of mediation analysis (Bullock et al., 2010; Green et al., 2010; Raykov & Marcoulides, 2001; Tomarken & Waller, 2003), the exploratory nature of our path modelling approach, and the reliance on 'rules of thumb' (Burnham et al., 2011; Burnham & Anderson, 2004; Raftery, 1995, 1999) and descriptive comparisons for model evaluation, we refrain from making strong claims about the causal nature of the indirect effects and encourage future

research employing confirmatory methods to investigate the underlying structure of effects in more detail.

5.5   Qualitative domain differences

We had set out to test possible domain differences in the context of quality of evidence. We explored the effects of availability of data (or lack thereof) versus expert agreement (scientific consensus) and expert disagreement (conflict). Stating that experts agree provides an explicit consensus signal (van der Linden et al., 2019). Stating that there is a lot of data available doesn't necessarily provide the same signal, as the various sources could disagree or not be of high quality. Knowing the powerful role scientific consensus plays in judgment and decision making on debated science (van der Linden et al., 2015, 2019), it could have been that, for the high quality-of-evidence condition, people would react more strongly to the 'consensus' messaging condition compared to the 'availability of data' condition. For the low quality-of-evidence conditions, research has shown that people respond negatively to consensus uncertainty in particular when compared to other kinds of uncertainty (Gustafson & Rice, 2020). However, our data do not allow us to adjudicate in favour or against any of these explanations, as we did not find significant differences between the groups that provided different explanations for the quality of the evidence level in any of our three studies, nor, in Study 3, between these conditions and providing people with a mere quality-of-evidence 'label' ('high' or 'low').

It may be that providing a label for quality of evidence that is as concise and strong as 'low' and 'high' constitutes such a strong signal that it outweighs all additional information, or it may be that our manipulations were not strong enough to tap into the relevant constructs in this context. It may also depend on the kind of information. For instance, the context used in the present study — information on the COVID-19 case fatality rate — might have been quite abstract to people and not self-relevant enough to trigger an interest in the underlying reasoning. It is conceivable that in more immediately self-relevant contexts, for instance when it comes to choice of compliance with behavioural mandates or medical decision making, people might care more and be more interested in the reason for a quality-of-evidence level. However, perhaps people care more about the actual quality level and less about the reason for the label. Future research could be carried out in different contexts, and provide more detailed information on the reasoning, such as providing two or more expert estimates to illustrate what expert (dis)agreement looks like.

5.6   Limitations and future directions

Our work is not without limitations. We did not use probability samples to draw inferences on a population level. However, we did base our samples on national quotas which places the quality of our datasets above typical convenience samples frequently used in psychology research (Kees et al., 2017). A second limitation is that we used UK data for all of our studies.

We therefore encourage replications with multi-country datasets to test the generalizability of observed effects across cultures. Of note is also the fact that the observed effect sizes for differences between control, low and ambiguous quality-of-evidence groups were small to medium ($d = 0.19$–0.36), with bigger effect sizes for differences between high and low and ambiguous quality-of-evidence groups ($d = 0.24$–0.50). Recent research on effect sizes in psychological research has argued that even small effects can be consequential and that effects of medium size do have explanatory power and can be of practical use (Funder & Ozer, 2019). We therefore believe that when considering quality-of-evidence communications on a population level, even small effects should be considered when making decisions about public communications. Unfortunately, our surveys could not collect true behavioural measures but rely on people's self-reports; the 'intention-action' gap therefore has to be kept in mind (Sheeran & Webb, 2016).

Notwithstanding these limitations, this set of studies is one of the first to explore how people make judgments and decisions about quality-of-evidence indicators in a public health context. Future work could test further aspects of quality-of-evidence communication; such as, for instance, the role of the source of the quality-of-evidence rating. It is conceivable that there could be important differences in how people react to quality-of-evidence information if the judgment on quality levels comes from, for instance, public health experts versus journalists or a credible versus unknown source. It would also be useful to test the generalizability of the presented quality-of-evidence relationships outside of high profile cases, such as the COVID-19 context used in this study, in addition to exploring potential moderating relationships, such as whether and how effects might vary by for instance political orientation, education, or science scepticism.

References

Aklin, M., & Urpelainen, J. (2014). Perceptions of scientific dissent undermine public support for environmental policy. *Environmental Science and Policy*, *38*, 173–177. https://doi.org/10.1016/j.envsci.2013.10.006.

Baillon, A., Cabantous, L., & Wakker, P. P. (2012). Aggregating imprecise or conflicting beliefs: An experimental investigation using modern ambiguity theories. *Journal of Risk and Uncertainty*, *44*(2), 115–147. https://doi.org/10.1007/s11166-012-9140-x.

Benjamin, D. M., & Budescu, D. V. (2018). The role of type and source of uncertainty on the processing of climate models projections. *Frontiers in Psychology*, *9*(MAR), 1–17. https://doi.org/10.3389/fpsyg.2018.00403.

Bentler, P. M. (1990). Comparative fit indexes in structural models. In *Psychological Bulletin* (Vol. 107, Issue 2, pp. 238–246). American Psychological Association. https://doi.org/10.1037/0033-2909.107.2.238.

Blastland, M., Freeman, A. L. J., Linden, S. Van Der, Marteau, T. M., & Spiegelhalter, D. (2020). Five rules for evidence communication. *Nature*, *587*(November), 362–364.

https://doi.org/10.1038/d41586-020-03189-1.

Brzezinski, A., Kecht, V., Van Dijcke, D., & Wright, A. L. (2021). Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States. *Nature Human Behaviour*, *5*(11), 1519–1527. https://doi.org/10.1038/s41562-021-01227-0.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology*, *98*(4), 550–558. https://doi.org/10.1037/a0018933.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644.

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. https://doi.org/10.1007/s00265-010-1029-6.

Cabantous, L. (2007). Ambiguity aversion in the field of insurance: Insurers' attitude to imprecise and conflicting probability estimates. *Theory and Decision*, *62*(3), 219–240. https://doi.org/10.1007/s11238-006-9015-1.

Cabantous, L., Hilton, D., Kunreuther, H., & Michel-Kerjan, E. (2011). Is imprecise knowledge better than conflicting expertise? Evidence from insurers' decisions in the United States. *Journal of Risk and Uncertainty*, *42*(3), 211–232. https://doi.org/10.1007/s11166-011-9117-1.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. https://doi.org/10.1007/BF00122575.

Collins, R. N., & Mandel, D. R. (2019). Cultivating credibility with probability words and numbers. *Judgment and Decision Making*, *14*(6), 683–695. https://doi.org/10.31234/osf.io/gkduw.

Dhami, M. K., & Mandel, D. R. (2021). Words or numbers? Communicating probability in intelligence analysis. *American Psychologist*, *76*(3), 549–560. https://doi.org/10.1037/amp0000637.

ECDC. (2020). *European Centre for Disease Control COVID-19 Situation Update*. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.

EFF. (2020). *Teaching and Learning Toolkit*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit.

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, *111*(Supplement_4), 13664–13671. https://doi.org/10.1073/pnas.1317504111.

Fox, C., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, *110*(3), 585–603. https://doi.org/10.2307/2946693.

Frisch, D., & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision*

Making, *1*(3), 149–157. https://doi.org/10.1002/bdm.3960010303.

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119. https://psycnet.apa.org/doi/10.1037/0003-066X.46.2.107.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221–233. https://doi.org/10.1037/0022-3514.65.2.221.

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "Black Box" experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Science*, *628*(1), 200–208. https://doi.org/10.1177/0002716209351526.

Griffin, D., & Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology*, *24*(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R.

Gustafson, A., & Rice, R. E. (2019). The Effects of Uncertainty Frames in Three Science Communication Topics. *Science Communication*, 107554701987081. https://doi.org/10.1177/1075547019870811.

Gustafson, A., & Rice, R. E. (2020). A review of the effects of uncertainty in public science communication. *Public Understanding of Science*, *29*(6), 614–633. https://doi.org/10.1177/0963662520942122.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*, 924–926. https://doi.org/10.1136/bmj.39489.470347.AD.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., & Schünemann, H. J. (2008). GRADE: what is "quality of evidence" and why is it important to clinicians? *BMJ*, *336*, 995–998. https://doi.org/10.1136/bmj.39490.551019.be.

Han, P. K. J. (2013). Conceptual, Methodological, and Ethical Problems in Communicating Uncertainty in Clinical Evidence. *Med Care Res Rev.*, *70*(10), 14S-36S. https://doi.org/10.1177/1077558712459361.

Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2011). Varieties of uncertainty in health care: A conceptual taxonomy. *Medical Decision Making*, 828–838. https://doi.org/DOI: 10.1177/0272989X11393976.

Hendriks, F., & Jucks, R. (2020). Does scientific uncertainty in news articles affect readers' trust and decision-making? *Media and Communication*, *8*(2), 401–412. https://doi.org/10.17645/mac.v8i2.2824.

Howe, L. C., MacInnis, B., Krosnick, J. A., Markowitz, E. M., & Socolow, R. (2019). Acknowledging uncertainty impacts public acceptance of climate scientists' predictions.

Nature Climate Change, *9*(11), 863–867. https://doi.org/10.1038/s41558-019-0587-5.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362. https://doi.org/10.1037/0033-2909.112.2.351.

IPCC. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In Core Writing Team, R. K. Pachauri, & L. A. Meyer (Eds.), *IPCC*. https://www.ipcc.ch/report/ar5/syr/.

Irwin, D., & Mandel, D. R. (2019). Improving information evaluation for intelligence production. *Intelligence and National Security*, *34*(4), 503–525. https://doi.org/10.1080/02684527.2019.1569343.

Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2017). Maintaining credibility when communicating uncertainty: The role of communication format. *Proceedings of the 39th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*, 582–587. https://discovery.ucl.ac.uk/id/eprint/10039458/.

Jenkins, S. C., & Harris, A. J. L. (2021). Maintaining credibility when communicating uncertainty: the role of directionality. *Thinking and Reasoning*, *27*(1), 97–123. https://doi.org/10.1080/13546783.2020.1723694.

Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2018). When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. *Journal of Risk Research*, *9877*, 1–18. https://doi.org/10.1080/13669877.2018.1440415.

Joslyn, S., & Demnitz, R. (2021). Explaining how long CO2 stays in the atmosphere: Does it change attitudes toward climate change? *Journal of Experimental Psychology: Applied*, *27*(3), 473–484. https://doi.org/10.1037/xap0000347.

Joslyn, S. L., & Leclerc, J. E. (2016). Climate Projections and Uncertainty Communication. *Topics in Cognitive Science*, *8*(1), 222–241. https://doi.org/10.1111/tops.12177.

Kause, A., Bruin, W. B., Domingos, S., Mittal, N., Lowe, J., & Fung, F. (2021). Communications about uncertainty in scientific climate-related findings: a qualitative systematic review. *Environmental Research Letters*, *16*(5). https://doi.org/10.1088/1748-9326/abb265.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. https://doi.org/10.1080/00913367.2016.1269304.

Keren, G., & Gerritsen, L. E. M. (1999). On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, *103*, 149–172. https://doi.org/10.1016/S0001-6918(99)00034-7.

Koehler, D. J., & Pennycook, G. (2019). How the public, and scientists, perceive advance-

ment of knowledge from conflicting study results. *Judgment and Decision Making*, *14*(6), 671–682. https://journal.sjdm.org/19/191002b/jdm191002b.pdf.

Kreps, S. E., & Kriner, D. L. (2020). Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic. *Science Advances*, *6*(43), eabd4563. https://doi.org/10.1126/sciadv.abd4563.

Lei, P.-W., & Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33–43. https://doi.org/10.1111/j.1745-3992.2007.00099.x.

Machina, M. J., & Siniscalchi, M. (2014). Ambiguity and ambiguity aversion. In *Handbook of the Economics of Risk and Uncertainty* (Vol. 1). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53685-3.00013-1.

Mandel, D. R., Wallsten, T. S., & Budescu, D. V. (2021). Numerically Bounded Linguistic Probability Schemes Are Unlikely to Communicate Uncertainty Effectively. *Earth's Future*, *9*(1), 1–5. https://doi.org/10.1029/2020EF001526.

Mandel, D. R., & Irwin, D. (2021). Facilitating sender-receiver agreement in communicated probabilities: Is it best to use words, numbers or both? *Judgment and Decision Making, 16*(2), 363–393. https://journal.sjdm.org/20/200930/jdm200930.html.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., & Zwiers, F. W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. *IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties, Jasper Ridge, CA, USA, 6–7 July 2010*. https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf.

McNeish, D. (2018). Should We Use F-Tests for Model Fit Instead of Chi-Square in Overidentified Structural Equation Models? *Organizational Research Methods*, *23*(3), 487–510. https://doi.org/10.1177/1094428118809495.

O'Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies*, *26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063.

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. https://doi.org/10.1177%2F0049124199027003005.

Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, *8*(1), 142–149. https://doi.org/10.1207/S15328007SEM0801_8.

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1–10. https://doi.org/10.1057/s41599-019-0279-9.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of*

881

Statistical Software, *48*(2 SE-Articles), 1–36. https://doi.org/10.18637/jss.v048.i02.

Schneider, C. R., Freeman, A. L. J., Spiegelhalter, D., & van der Linden, S. (2021). The effects of quality of evidence communication on perception of public health information about COVID-19: Two randomised controlled trials. *Plos One*, *16*(11), e0259048. https://doi.org/10.1371/journal.pone.0259048.

Sheeran, P., & Webb, T. L. (2016). The intention-behavior gap. *Socialurl and Personality Psychology Compass*, *10*(9), 503–518. http://eprints.whiterose.ac.uk/107519/.

Smithson, M. (1999). Conflict aversion: Preference for ambiguity vs conflict in sources and evidence. *Organizational Behavior and Human Decision Processes*, *79*(3), 179–198. https://doi.org/10.1006/obhd.1999.2844.

Smithson, M. (2015). Probability judgments under ambiguity and conflict. *Frontiers in Psychology*, *6*(May), 1–9. https://doi.org/10.3389/fpsyg.2015.00674.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578–598. https://psycnet.apa.org/doi/10.1037/0021-843X.112.4.578.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

UK Government. (2020). *Scientific Advisory Group for Emergencies (SAGE)*. https://www.gov.uk/government/organisations/scientific-advisory-group-for-emergencies.

van der Bles, A. M., van der Linden, S., Freeman, A. L. J. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(14), 7672–7683. https://doi.org/10.1073/pnas.1913678117.

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, *6*(5). https://doi.org/10.1098/rsos.181870.

van der Linden, S., Clarke, C., & Maibach, E. (2015). Highlighting consensus among medical scientists increases public support for vaccines: evidence from a randomized experiment. *BMC Public Health*, *15*(1207). https://doi.org/10.1186/s12889-015-2541-4.

van der Linden, S., Leiserowitz, A., & Maibach, E. (2019). The gateway belief model: A large-scale replication. *Journal of Environmental Psychology*, *62*, 49–58. https://doi.org/10.1016/j.jenvp.2019.01.009.

Veit, W., Brown, R., & Earp, B. D. (2021). In Science We Trust? Being Honest About the Limits of Medical Research During COVID-19. *American Journal of Bioethics*, *21*(1), 22–24. https://doi.org/10.1080/15265161.2020.1845861.

Weiss, C. (2003). Expressing scientific uncertainty. *Law, Probability, and Risk*, *2*, 25–46. https://doi.org/10.1093/lpr/2.1.25.

882