The Einstein effect: Global evidence for scientific source credibility effects and the influence of religiosity

Suzanne Hoogeveen*[1], Julia M. Haaf[1], Joseph A. Bulbulia[2], Robert M. Ross[3], Ryan McKay[4], Sacha Altay[5], Theiss Bendixen[6], Renatas Berniūnas[7], Arik Cheshin[8], Claudio Gentili[9], Raluca Georgescu[10], Will M. Gervais[11], Kristin Hagel[12], Christopher Kavanagh[13,14], Neil Levy[3], Alejandra Neely[15], Lin Qiu[16], André Rabelo[17], Jonathan E. Ramsay[18], Bastiaan T. Rutjens[1], Hugh Turpin[13], Filip Uzarevic[19], Robin Wuyts[1], Dimitris Xygalatas[20], and Michiel van Elk[21]

[1]University of Amsterdam
[2]Victoria University of Wellington
[3]Macquarie University
[4]Royal Holloway, University of London
[5]Institut Jean Nicod
[6]Aarhus University
[7]Vilnius University
[8]University of Haifa
[9]University of Padova
[10]Babes-Bolyai University
[11]Brunel University London
[12]Max Planck Institute for Evolutionary Anthropology
[13]University of Oxford
[14]Rikkyo University
[15]Adolfo Ibáñez University
[16]Nanyang Technological University
[17]Universidade de Brasília
[18]James Cook University
[19]Catholic University of Louvain
[20]University of Connecticut
[21]Leiden University

*Correspondence should be sent to Suzanne Hoogeveen, University of Amsterdam, Nieuwe Achtergracht 129 B, 1018 WT Amsterdam, The Netherlands. E-mail may be sent to suzanne.j.hoogeveen@gmail.com.
Data, analysis code, and stimuli are provided at https://osf.io/qsyvw/. Full materials in each language can be found at osf.io/kywjs/.

Abstract

People tend to evaluate information from reliable sources more favourably, but it is unclear exactly how perceivers' worldviews interact with this source credibility effect. In a large and diverse cross-cultural sample ($N = 10,195$ from 24 countries), we presented participants with obscure, meaningless statements attributed to either a spiritual guru or a scientist. We found a robust global source credibility effect for scientific authorities, which we dub 'the Einstein effect': across all 24 countries scientists hold greater authority than spiritual source, even among highly committed religious people, who are relatively also more credulous of nonsense from scientists than they are of nonsense from spiritual gurus. Additionally, individual religiosity predicted a weaker relative preference for the statement from the scientist vs. the spiritual guru, and was more strongly associated with credibility judgments for the guru than the scientist. Independent data on explicit trust ratings across 143 countries mirrored the experimental patterns. These findings suggest that irrespective of religious worldview, science is a powerful and universal heuristic that signals the reliability of information.

Keywords: source credibility, religion, science beliefs, culture

In a heated debate about the proximity of COVID-19 herd immunity, White House health advisor Dr. Scott Atlas proclaimed "You're supposed to believe the science, and I'm telling you the science"[1]. A group of infectious disease experts and former colleagues from Stanford, however, publicly criticized Dr. Atlas, who is a radiologist, for spreading 'falsehoods and misrepresentation of science' through his statements about face masks, social distancing and the safety of community transmission[2]. In the 2020 pandemic crisis, all eyes turned to scientific experts to provide advice, guidelines and remedies; from COVID-19 alarmists to skeptics, appeal to scientific authority appeared a prevalent strategy on both sides of the political spectrum. Please see the Appendix for a short commentary on how the present work might relate to the COVID-19 situation.

A large body of research has shown that the credibility of a statement is heavily influenced by the perceived credibility of its source[3–10]. Children and adults are sensitive to the past track record of informants[11–16], evidence of their benevolence toward the recipient of testimony[17–19], as well as how credible the information is on its face[20,21]. From an evolutionary perspective, deference to credible authorities such as teachers, doctors, and scientists is an adaptive strategy that enables effective cultural learning and knowledge transmission[22–28]. Indeed, if the source is considered a trusted expert, people are willing to believe claims from that source without fully understanding them. We dub this 'the Einstein effect'; people simply accept that $E = mc^2$ and that antibiotics can help cure pneumonia because credible authorities such as Einstein and their doctor say so, without actually understanding what these statements truly entail.

Knowing that a statement originates from an epistemic authority may thus increase the likelihood of opaque messages being interpreted as meaningful and profound. According to Sperber[29], in some cases, incomprehensible statements from credible sources may be appreciated not just in spite of but *by virtue of* their incomprehensibility, as exemplified by the speech of spiritual or intellectual gurus (the "Guru effect"). Here, we investigate to what extent different epistemic authorities affect the perceived value of nonsensical information. To this end, we contrasted judgements of gobbledegook spoken by a spiritual leader with gobbledegook spoken by a scientist. In addition, we assessed whether the source effect is predicted by individual religiosity and varies cross-culturally, as a proxy for how scientists and spiritual authorities function as "gurus" for different individuals and within different cultural contexts.

Although source credibility effects have typically been investigated for persuasion in marketing and communication, both science and spirituality may present particularly suitable contexts for inducing strong source effects. Scientists are generally considered competent and benevolent sources[30,31] and scientific information is often difficult and counterintuitive[32–34]. The combination of a credible authority and intangible information can increase the probability of obscure scientific information being accepted, by enhancing perceivers' reliance on the source[9,10,35]. Even indirect context cues, such as those emphasizing the scientific nature of a piece of information can increase the probability that (dubious) information is believed[36]. Some experimental evidence, for instance, suggests that irrelevant neuroscience information[37–39] or nonsense mathematical equations[40] can boost the perceived quality of presented claims, though note that replication studies suggest that mere brain images may not suffice[41,42]. Notably, these effects were only present among nonexperts (i.e., people with little formal neuroscientific or mathematical training). This distinction suggests that the appeal of "sciencey" information may be particularly strong when analytical assessment fails and one can only rely on secondary credibility cues.

Similar to the anticipated complexity of scientific information, prior beliefs about religious or spiritual texts instigate expectations that the information presented will be obscure. Supernatural explanations often appeal to phenomena that operate outside of the natural world and to experiences deemed ineffable, mysterious and exempt from empirical validation[43–48]. Some scholars have argued that incomprehensible theological language and irrational beliefs may serve as a costly signal towards the religious ingroup, signalling quality by hard-to-fake moral commitment, intellectual capacity and epistemological investment[49,50]. However, irrespective of content biases, the evaluation of spiritual or theological obscurity critically depends on one's personal beliefs about the credibility of spiritual gurus or religious authorities.

Various lines of evidence suggest that perceived credibility of both content and source indeed depends on individual difference factors such as the perceiver's (political) ideology and worldview[51–54]. In the absence of the means to rationally evaluate a claim and reliable source information, people likely infer credibility based on beliefs about the group to which the source belongs (e.g., 'conservatives', 'scientists'). In this process, similarities between one's own worldview and that of the source's group may serve as a proxy for being a benevolent and reliable source[23,55]. In a religious context, Christians were found to be more affected by an intercessory prayer when supposedly performed by a (charismatic) Christian than a non-Christian[56] and to require less evidence for religious claims (e.g., efficacy of prayer to cure illness) than for scientific claims (e.g., efficacy of medication[57,58]). These differences were not present among secular individuals. Furthermore, evangelical Christians were more likely to accept statements opposing their personal views when attributed to an ingroup religious leader versus an outgroup religious leader[59]. This effect was moderated by the amount of contact participants had with the specific group the religious leader belonged to, which highlights the importance of the person-source fit for message acceptance.

To account for these effects, alongside traditional dual-process models of persuasion[9,10,60,61], various authors have recently proposed a Bayesian framework in which subjective beliefs about the source (e.g., trustworthiness) and one's worldviews contribute to belief updating in response to new information following Bayesian principles[6,62–64]. By including background beliefs, these Bayesian networks describe how a differential weighing of evidence and even divergent updating (belief polarization) can be considered rational and normative. This may explain, for instance, how strong religious believers can become more convinced of their beliefs in the face of disconfirmatory evidence, especially when their faith is being challenged[63,65]. Similarly, strong conservatives who distrust science may become less convinced of human-caused global warming when presented with scientific consensus information[62]. In other words, laypeople may apply their own 'power priors'[66] to calibrate evidence from different sources, whose trustworthiness is subjectively determined, partly by their broader worldview.

In sum, whereas previous studies have established source credibility effects in a wide array of domains, as-of-yet little is known about whether and to what extent people's worldview is predictive of the relative credibility evaluation of information from scientific and spiritual sources. In the present study, we presented participants ($N = 10,195$, from 24 countries) with meaningless verbiage (henceforth, "gobbledegook"; also referred to in the literature as "pseudo-profound bullshit"[67]) randomly credited to either a spiritual authority or a scientific authority (see Figure 5). We assessed (1) whether trusting scientific experts over spiritual leaders is a general heuristic (i.e., the Einstein effect), and (2) to what extent perceivers' religiosity predicts the relative confidence in the truth of the gobbledegook statements from both sources. Note that we chose a "spiritual guru" authority frame, instead of "religious leader," because we wanted to avoid selecting an authority specific to any particular religion, to keep the study consistent across countries. While religiosity and spirituality are overlapping but not interchangeable constructs[68,69], self-reported religiosity has been positively associated with belief in spiritual phenomena such as fate, spiritual energy, and a connected universe[70–72] (though not unequivocally[73]). Consequently, we expected religiosity to be associated with increased receptivity to gobbledegook from a spiritual authority.

All confirmatory hypotheses and included measures were preregistered on the Open Science Framework (see osf.io/faj2z/). This link contains the original preregistration file. The registered component (including additional sub-projects) can be found at osf.io/xg8y5/files. In addition, for exploratory purposes, we included response time measures and a memory test to obtain insight into the cognitive processes underlying the source credibility effect (these measures were anticipated in the preregistration, but no concrete hypotheses were formulated). In order to further validate the findings from our experimental paradigm, we also analysed a large dataset from 117,191 individuals across 143 countries (including the same countries included in our study) that contains explicit trust ratings of scientists and traditional healers, as well as participant religiosity[74].

Results

The two dependent variables that were measured (i.e., *importance* of the message and *credibility* of the message) were highly correlated for both the scientific source (Spearman's $\rho = 0.772$, 95% credible interval $[0.764, 0.779]$) and for the spiritual source (Spearman's $\rho = 0.827$, 95% credible interval $[0.822, 0.833]$; see Figure A7)[75]. As the pattern of results was equal across the

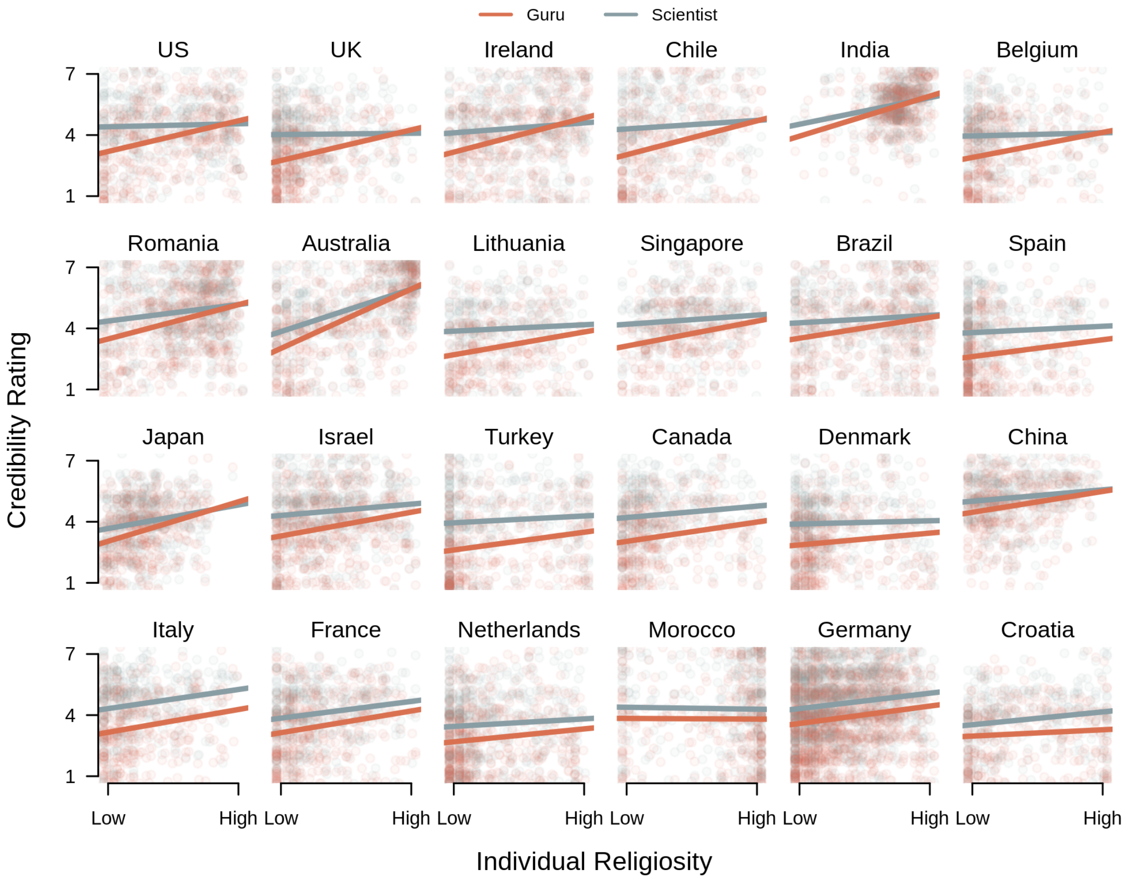


Figure 1: Observed relation between religiosity and credibility ratings per source, for each country. Countries are ordered by size of the source-by-religiosity interaction (from left to right, top to bottom). Red lines denote ratings for the spiritual guru and grey lines denote ratings for the scientist. Data points are jittered to enhance visibility. Credibility was measured on a 7-point Likert scale.

dependent variables, we decided to only describe the findings for *credibility* in detail (see Table 2 for the results for *importance*).

Effect of source on credibility

First, we assessed the extent to which the perceived credibility of a gobbledegook statement is affected by its source (i.e., a scientist vs. a spiritual guru). Note, our initial hypothesis was that there would be no main effect of source, that is, we expected evidence for the null-model. However, based on visual inspection of the data (see Figure 1) a main effect of source seems evident. To quantify the evidence for the effect of source, we compared between the null model without an effect of condition (i.e., the scientist and spiritual guru are judged equally credible), the model with a *common positive effect* of condition across countries (i.e., the scientist is judged more credible than the guru, to an equal degree in every country), the model with a *varying positive effect* of source (i.e., the scientist is judged more credible than the guru, but to varying degrees across countries), and the *unconstrained* model that allows the source effect to be varying from both positive to negative (i.e., in some countries, the scientist is considered more credible than the guru, in other countries, the guru is considered more credible than the scientist).

The Bayes factor model-comparison summarized in Table 1 shows that the data provide

Table 1: Bayes factor model comparisons to test $\mathcal{H}_1$ and $\mathcal{H}_2$

Model		Bayes factor	$p(\mathcal{M})$
Hypothesis 1: Source effect			
$\mathcal{M}_0$	$\text{Country}_u + \text{Religiosity}_u$	1-to-10^{228}	$< .01$
$\mathcal{M}_1$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_1$	1-to-10^{17}	$< .01$
$\mathcal{M}_+$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_+$	$*$	.92
$\mathcal{M}_u$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_u$	1-to-12.30	.08
Hypothesis 2: Source-by-Religiosity Effect			
$\mathcal{M}_0$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_u$	1-to-10^{15}	$< .01$
$\mathcal{M}_1$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_u + \text{Source*Religiosity}_1$	$*$	.50
$\mathcal{M}_+$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_u + \text{Source*Religiosity}_+$	1-to-1.28	.39
$\mathcal{M}_u$	$\text{Country}_u + \text{Religiosity}_u + \text{Source}_u + \text{Source*Religiosity}_u$	1-to-4.60	.11

Note. Asterisks mark the preferred model for each hypothesis. The remaining values are the Bayes factors for the respective model vs. the preferred model. Subscripts reflect parameter constraints; $_u$ indicates an unconstrained effect, $_1$ indicates a common (positive/negative) effect, $_+$ indicates a varying positive/negative effect. $p(\mathcal{M})$ gives the posterior model probability per hypothesis. All models include the covariate level of education.

most evidence for the *positive effects model*, which assumes a varying but consistently positive effect across countries. The source effect is favoured 1.1×10^{210}-to-1 over the null-model, which indicates strong evidence that the meaningless statement from the scientist is considered more credible than the meaningless statement from the guru. The positive effects model strongly outperforms the common effect model ($\text{BF}_{+1} = 8.9 \times 10^{17}$; explained variance (Bayesian R^2) is 17.9%, 95% credible interval [17.0%, 18.7%]). The mean and 95% credible interval of the unstandardized size of the source effect in the full model is 0.70 [0.60, 0.79] on a 7-point Likert scale and the standard deviation between countries is 0.16. Also note that as shown in Figure 1 the within-country individual differences in credibility ratings are large, indicating that most of the variance is located at the lower level (i.e., the individual level). The intraclass correlation coefficients (ICCs) quantifying the proportion of variance explained by the country clustering, as well as the total explained variance by the included effects for all models (Bayesian R^2) are reported in the Appendix. There, we also report MCMC diagnostics to verify the adequacy of the Bayesian models, as well as the estimates for the intercepts, source effect, and the source-by-religiosity interaction effect for each country.

The fit-effect: Interaction between source and religiosity on credibility

The *source-by-religiosity interaction effect* assesses to what extent the effect of source depends on raters' own religious background (religiosity was globally standardised). Our hypothesis states that for low religious individuals, credibility ratings should be higher for gobbledegook from a scientific source than for gobbledegook from a spiritual guru. For highly religious individuals, the reversed effect is expected, i.e., higher credibility ratings for gobbledegook ascribed to a guru than for gobbledegook ascribed to a scientist. The interaction term was therefore constrained to be *negative*, in the sense that the coefficient of the source effect becomes smaller (or negative) with increased religiosity. Note that although the interaction term was constrained to have a negative sign, for consistency, we still refer to the model as the positive effects model.

For hypothesis 2, the model comparison summarized in Table 1 shows that the data provide most evidence for the *common source-by-religiosity interaction model*, which assumes a consistent interaction effect across countries, $\text{BF}_{10} = 0.99 \times 10^{15}$ ($R^2 = 18.1\%$ [17.2%, 19.0%]). The data are uninformative for distinguishing between the common interaction and the varying positive interaction model ($\text{BF}_{1p} = 1.28$), indicating that both are equally plausible. While we cannot conclude whether or not the size of the interaction effect differs substantially between countries, both models provide strong evidence for a source-by-religiosity effect across all countries. The mean of the unstandardized source-by-religiosity interaction effect is -0.21 [-0.29, -0.14] and the standard deviation between countries is 0.09 on the 7-point Likert scale. As becomes evident from Figure 2d, the interaction entails that the relative preference in credibility for statements from the scientist versus the spiritual guru decreases with higher religiosity. This effect is further unpacked in Figure 2c, which shows that in every country, except for Croatia, religiosity is more predictive of credibility ratings for statements from the guru than for statements from the scientist.

Exploratory Analyses

In an exploratory fashion, we assessed to what extent the source manipulation influenced the effort participants put into processing the statements. To this end, we looked at (1) response time for the evaluation of each statement as a proxy for processing time of the message, and (2) memory performance of words presented in the statements as a proxy for encoding quality. For these exploratory models, we only assessed evidence for a common effect, as visual inspection of the data suggested no or only very small and homogeneous effects (see Figure 3).

Processing Time

For processing time the data indicate a common effect of source: participants spent more time processing the statement of the scientist (median RT = 28.30 seconds) than that of the guru (median RT = 27.0 seconds; BF_{10} = 8,050.48). Processing times were log-transformed for the analysis, to account for the positive skew that is typically observed in response time data. However, the standardized effect size is very small: 0.058 [0.023, 0.087]. There was strong evidence against an interaction between source and religiosity ratings on processing time: religiosity is not predictive of the difference in processing time for the scientist vs. the guru ($\text{BF}_{10} = 0.03$, $\text{BF}_{01} = 30.78$).

Memory Performance

After the rating question, participants were presented with a recall item that required them to indicate which words they recognized from the statement. The list consisted of 5 target (included in the statement) and 5 distractor words (not in the statement) for each source. An F_1 score was calculated per person per source, which gives the harmonic mean of the precision (proportion true positives of all selected words) and recall (proportion true positives of all presented target words). F_1 ranges between 0 and 1, with 1 being perfect performance.

The analysis indicated some evidence against a common effect of source on memory performance: participants did not perform better on recognising words from the statement by the scientist than by the guru ($\text{BF}_{10} = 0.53$; $\text{BF}_{01} = 1.90$; standardized estimate = 0.014 [0.001, 0.035]). Finally, there was some evidence against an interaction, $\text{BF}_{10} = 0.31$, $\text{BF}_{01} = 3.27$.

As a sanity check, we showed that there is an extremely strong effect of processing time on memory performance; participants who spent more time processing the statement, also performed better on the memory task ($\text{BF}_{10} = \infty$).

Validation using previously collected trust ratings

In addition to the experimental data collected in this study, we also examined an existing dataset that includes surveyed trust ratings for scientists and traditional healers for 117,191 participants across 143 countries. Note that the analysis on this dataset was not preregistered. Analysis of these data corroborated the results from our experimental manipulations; on average scientists are considered more trustworthy than traditional healers, standardized estimate = 0.30 [0.06, 0.58] (for comparison: the standardized estimate for the experimental source effect on credibility is 0.41 [0.22, 0.49]). While the positive effects model strongly outperforms both the null model and the common effect model (BF_{+0}, $\text{BF}_{+1} > 10^{308}$; R^2 for the positive effects model $= 28.1\% \, [27.8\%, 28.3\%]$), the analysis indicates most evidence for the unconstrained model $\mathcal{M}_u$, which indicates that scientists are not explicitly trusted more than traditional healers in all of the 143 countries, $\text{BF}_{u+} = 320.76$. Nonetheless, as displayed in Figure 4a, only in 3 out of the 143 countries the mean of the estimated source effect is negative, while the overall effect is clearly positive.

We also investigated the fit-effect in this dataset, by including an interaction term between authority (scientists vs. traditional healers) and religiosity (religious vs. not religious). Because in 41 countries all of the participants indicated that they were religious, we could not reliably estimate varying effects for the authority-by-religiosity interaction. There was, however, strong evidence for an overall interaction between authority and religiosity, $\text{BF}_{10} = 6.3 \times 10^{14}$, $R^2 = 28.1\% \, [27.8\%, 28.4\%]$ standardized estimate = -0.09 [-0.14, -0.02] (for comparison: the standardized estimate for the experimental source-by-religiosity effect on credibility is -0.12 [-0.16, -0.08]). The pattern of the interaction is the same as for the experimental credibility data: the relative difference between trust in scientists vs. traditional healers is smaller for religious individuals than for non-religious individuals. Interestingly, while the experimental study found that religiosity was associated with increased credibility ratings for both sources, albeit to a smaller extent for the scientist (see Figure 2c), the trust data show a positive effect of religiosity on trust for traditional healers (standardized estimate = 0.03 [0.02, 0.04]), yet a negative effect of religiosity on trust for scientists (standardized estimate = -0.01 [-0.02, -0.01]). See the Appendix for an additional exploratory analysis on the country-level correlation in the source effect between the primary experimental dataset and secondary validation dataset on trust.

Robustness and additional checks

We conducted 8 additional analyses that the results should be robust against, including all specifications mentioned in the preregistration:

1. Excluding observations for which participants did not correctly recall the source of the statement ($n_{obs} = 1616$ [7.95%]);
2. Excluding data from Lithuania because $n < 300$ (as preregistered);
3. Using a different, less informed prior setting for r scale; $r = \frac{\sqrt{2}}{2} \approx 0.707$, corresponding to a 'wide' prior scale provided in the `BayesFactor` package[76];
4. Using the *importance* rating instead of the *credibility* rating as the outcome variable.
5. Applying a between-subjects design by only taking the first observation per participant.
6. Including all participants, including those who failed the attention check.
7. Running the analyses without adding any predictors as covariates;
8. Running the analyses including all covariates that might affect either the independent variable (religiosity) or the dependent variable (credibility ratings): statement version (A or B), presentation order (guru–scientist or scientist–guru), participant age (in decades), participant gender, level of education, and perceived socio-economic status (SES).

The results of these robustness analyses are given in Table 2 and corroborate the conclusions from the main analyses: the data indicate (a) a source effect that varies between countries but is consistently positive (scientist > guru), and (b) a positive source-by-religiosity interaction effect (either a common or varying effect).

Table 2: Bayes factor of different models for robustness checks

Robustness Set	N_{obs}	Estimate [95%CI]	BF_{10}	BF_{+1}	Preferred
Source effect					
Main analysis	20,318	0.70 [0.60, 0.79]	10^{210}	10^{17}	$\mathcal{M}_+$
Excluding source incorrect	18,702	0.78 [0.69, 0.88]	10^{249}	10^{15}	$\mathcal{M}_+$
Excluding Lithuania ($n < 300$)	19,736	0.69 [0.59, 0.79]	10^{200}	10^{17}	$\mathcal{M}_+$
Default prior settings	20,318	0.70 [0.56, 0.84]	10^{210}	10^{15}	$\mathcal{M}_+$
Importance as outcome variable	20,318	0.53 [0.43, 0.63]	10^{113}	10^{11}	$\mathcal{M}_+$
Between-subjects design	10,159	0.83 [0.68, 0.98]	10^{145}	10^{20}	$\mathcal{M}_+$
Including all subjects	20,980	0.69 [0.59, 0.78]	10^{210}	10^{20}	$\mathcal{M}_+$
No covariates	20,318	0.70 [0.60, 0.79]	10^{199}	10^{17}	$\mathcal{M}_+$
All covariates	20,318	0.70 [0.60, 0.79]	10^{211}	10^{17}	$\mathcal{M}_+$
Fit Effect (Source*Religiosity)					
Main analysis	20,318	-0.21 [-0.29, -0.14]	10^{15}	0.78	$\mathcal{M}_1$
Excluding source incorrect	18,702	-0.23 [-0.32, -0.15]	10^{17}	4.85	$\mathcal{M}_+$
Excluding Lithuania ($n < 300$)	19,736	-0.21 [-0.29, -0.13]	10^{14}	0.90	$\mathcal{M}_1$
Default prior settings	20,318	-0.21 [-0.34, -0.09]	10^{13}	10^{-6}	$\mathcal{M}_1$
Importance as outcome variable	20,318	-0.18 [-0.26, -0.10]	10^{9}	0.02	$\mathcal{M}_1$
Between-subjects design	10,159	-0.22 [-0.33, -0.12]	10^{7}	4.67	$\mathcal{M}_u$
Including all subjects	20,980	-0.22 [-0.29, -0.14]	10^{15}	0.56	$\mathcal{M}_1$
No covariates	20,318	-0.22 [-0.29, -0.14]	10^{14}	0.77	$\mathcal{M}_1$
All covariates	20,318	-0.21 [-0.29, -0.13]	10^{16}	0.09	$\mathcal{M}_1$

Note. Across all eight sets of robustness checks, the results are qualitatively equal to those of the main analyses (column 1); the data indicate (a) a strong source effect that varies between countries but is consistently positive (scientist > guru), (b) a source-by-religiosity interaction effect (either a common or varying effect). Subscripts reflect parameter constraints; $_0$ indicates the null model, $_+$ indicates a varying positive effect, and $_1$ indicates a common effect. Preferred refers to the best predicting model based on the data.

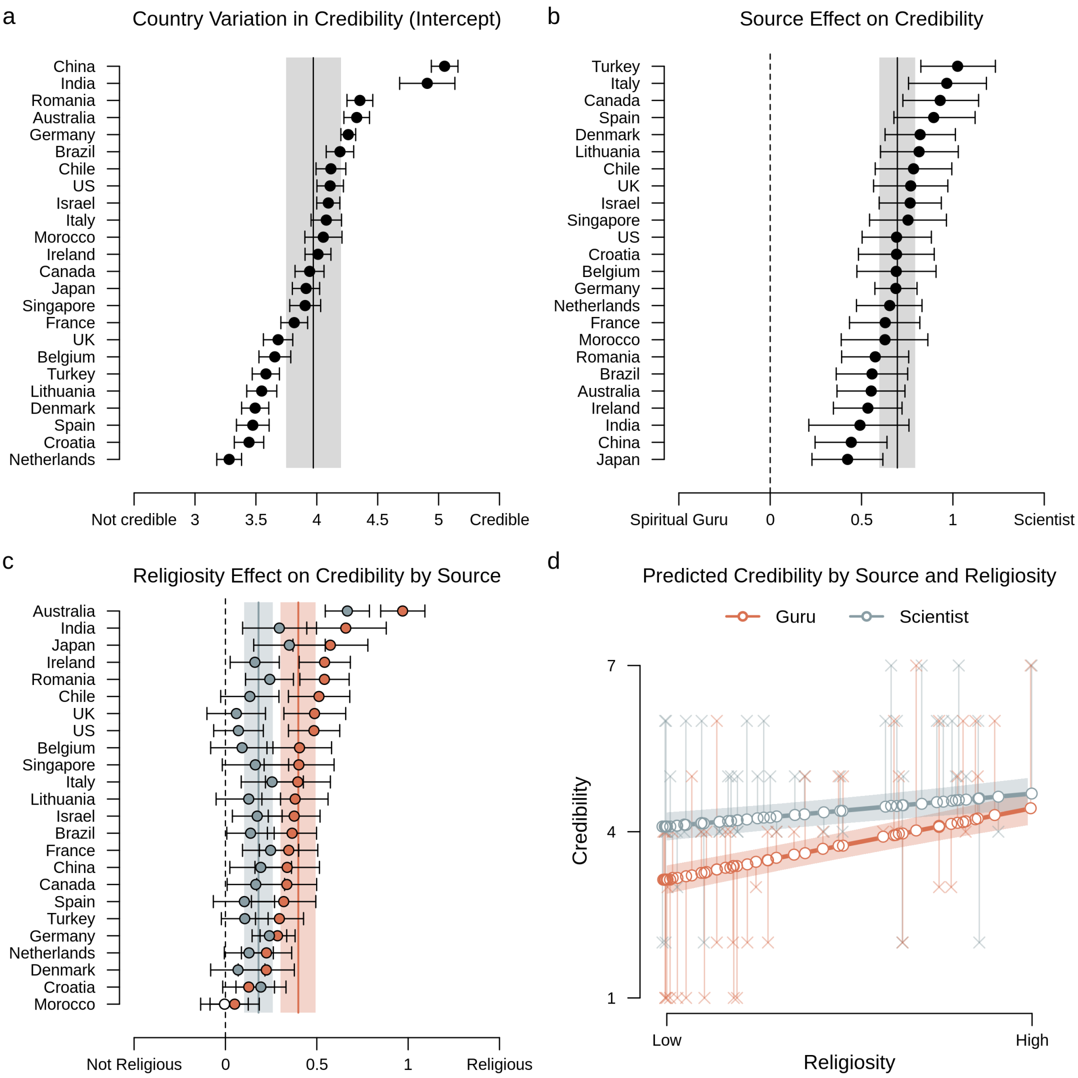


Figure 2: Summary of the multilevel-model (unconstrained) estimates per country and predicted overall effects. It is apparent that there is substantial variation across the 24 countries in (**a**) overall credibility judgments (i.e., intercept) and (**b**) the effect of scientific vs. spiritual source. Panel **c** shows that individual religiosity has a stronger effect on credibility judgments for the spiritual guru (red circles) than for the scientist (grey circles). The estimates are ordered from largest to smallest, and the open circles denote negatively valued effects. The errorbars give the 95% credible interval for each country. The vertical lines denote the overall estimated effect with the 95% credible interval in the shaded bands. The dashed lines indicates zero. Panel **d** displays the predicted credibility as a function of source and individual religiosity, showing that the difference in credibility ratings for the scientist (grey lines) vs. the guru (red lines) is less pronounced for high religiosity individuals than low religiosity individuals. The shaded bands reflects the 95% credible intervals, the x's reflect the observed values for 2 randomly sampled participants per country, and the circles reflect the corresponding estimated values. The x's and circles are jittered to enhance visibility.

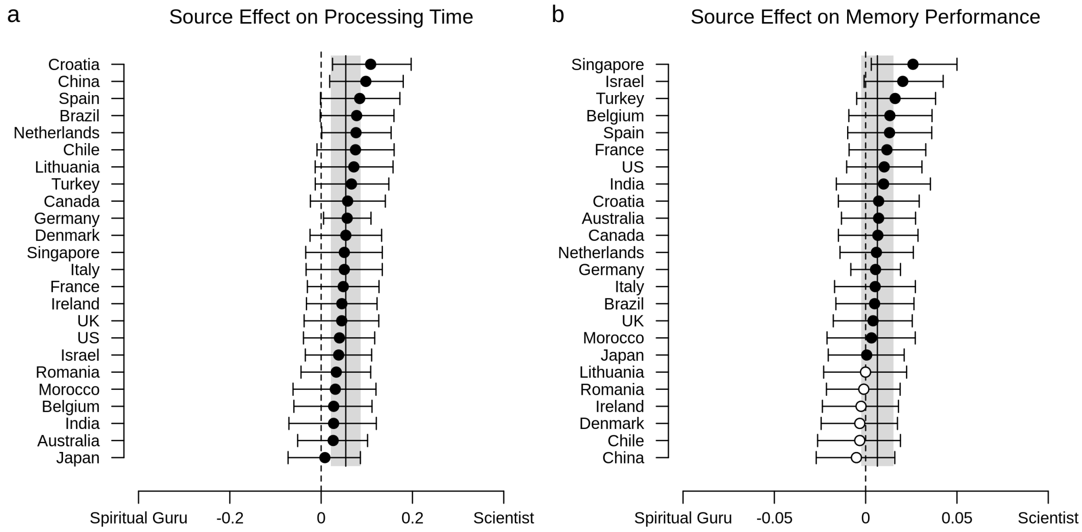


Figure 3: Multilevel-model (unconstrained) estimates for the source effect (**a**) on (log-transformed) processing time and (**b**) on memory performance (range 0–1). The estimates are ordered from largest to smallest, and the open circles denote negatively valued effects. The errorbars give the 95% credible interval for each country. The vertical lines denote the overall estimated effect with the 95% credible interval in the shaded bands. The dashed lines indicates zero.

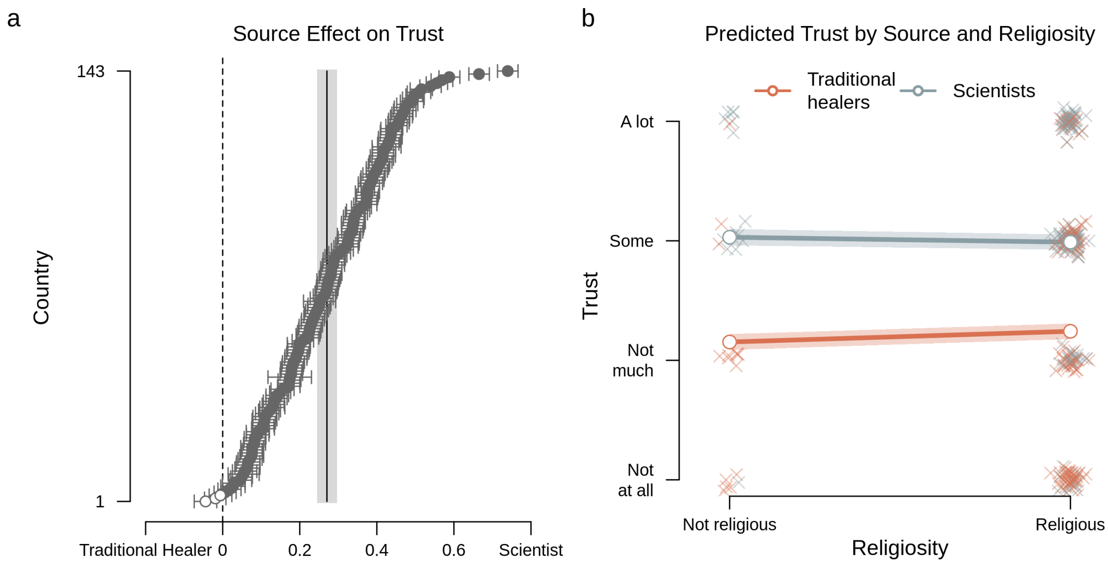


Figure 4: Multilevel-model (unconstrained) estimates and predicted overall effects for explicit trust ratings. Panel **a** displays the source effect on trust ratings for each of the 143 countries, showing that in all but 3 countries, scientists are trusted more than traditional healers. The estimates are ordered from largest to smallest, and the open circles denote negatively valued effects. The errorbars give the 95% credible interval for each country. The vertical lines denote the overall estimated effect with the 95% credible interval in the shaded bands. The dashed lines indicates zero. Panel **b** displays the predicted trust rating as a function of source and individual religiosity, showing that religious individuals trust scientists slightly less and traditional healers more compared to non-religious individuals. The shaded bands reflects the 95% credible intervals, the x's reflect the observed values for 2 randomly sampled participants per country, and the circles reflect the estimated values per condition. The x's are jittered to enhance visibility.

Discussion

In the current cross-cultural study, we used a straightforward manipulation and measurement of source credibility effects at the individual level. We found a robust source effect on credibility judgments of meaningless statements ascribed to different authority figures; across all 24 countries and all levels of religiosity, gobbledegook from a scientist was considered more credible than the same gobbledegook from a spiritual guru. In addition to this robust overall Einstein effect, participants' background beliefs predicted the credibility evaluations; individuals scoring low on religiosity considered the statement from the guru less credible than the statement from the scientist, while this difference was less pronounced for highly religious individuals. These patterns were consistent with explicit trust data collected for over 100,000 individuals from 143 countries: across 140 out of 143 of these countries, people indicated greater trust in scientists than in traditional healers, with a larger difference for non-religious compared to religious individuals. Robustness analyses for the experimental study indicated that the effects were robust against different data inclusion criteria (e.g., attention checks) and analytic choices (e.g., selection of covariates, dependent variable, prior settings). Moreover, the effects also compellingly emerged when analysed as a between-subjects design (see Table 2), suggesting that they are not simply explained by social desirability or participants responding in line with their guess of the research hypothesis (also note that recent empirical work indicates that online survey experiments are generally robust to experimenter demand effects[77]). Results of exploratory reaction time analyses suggest that in addition to giving more positive evaluations, people may actually put more effort into processing information from credible sources (though they did not recall it better). In particular, participants spent more time and may have tried relatively harder to decipher the gobbledegook from the scientist, whereas prior scepticism may have steered some to immediately dismiss the information from the guru as nonsense.

The pattern of results suggests that variability in the source effect between individuals and countries is more strongly driven by differences in credibility of the spiritual authority than the scientific authority. Based on the literature one could consider various plausible hypotheses explaining cross-cultural variation in the source effects, for instance in terms of cultural religiosity, vertically vs. horizontally structured societies, general trust in authorities, and specific trust patterns toward religious and secular authorities[78–83]. However, while our analysis indicated quantitative differences in the size of the source effect between countries (i.e., varying positive effects), we did not find qualitative differences (i.e., changes in the direction or presence of the effect). Descriptively, the weakest source effects (i.e., smallest difference between the scientific and the spiritual source) are observed in Asian countries (Japan, China, India), possibly because the spiritual guru as presented in the survey more closely fits Eastern belief systems than Abrahamic traditions. However, this explanation remains speculative and we are hesitant to over-interpret the cross-national variability both in the overall credibility judgments and the effect of source. While we included main effects of age, gender, level of education and socio-economic status in the analyses, the different sampling strategies that were applied between countries also calls for caution in making inferences based on direct comparisons.

Our findings could reflect a universal gullibility with regard to gobbledegook statements: only a small minority of participants, regardless of their national or religious background, displayed candid scepticism towards the nonsense statements, and 76% of participants rated the scientist's gobbledegook at or above the midpoint of the credibility scale (vs. 55% for the guru). However, the notion of a general gullibility underlying the observed effects is not entirely supported by the data. The median response was the midpoint of the credibility scale. Participants may have primarily used the midpoint of the scale to indicate that they were uncertain about whether or not the claim was credible, i.e., to refrain from passing judgment at all[84–86]. This response might appear as a lack in motivation to critically reflect on the information that was presented; at the same time, saving one's cognitive resources can also be considered 'strategic'. First, as with most psychology experiments, our study was a zero-stakes task with no incentive for accuracy, which may have lowered effort and biased responses toward the midpoint. Second, when analytic reasoning about the plausibility of a presented claim does not yield any conclusion, the most rational thing to do may be either suspending judgment (selecting the neutral midpoint of the rating scale) or calibrating judgment to prior beliefs about the source of the claim. If one considers the group to which the source belongs generally competent and benevolent, it makes sense to give a positive judgment of their difficult-to-evaluate claim. After all, credible experts often acquired credentials based on their reputation of discovering phenomena that seem implausible at first glance[55]. For instance, the premises of using vaccines ('inserting a virus prevents disease') or facts about climate change ('humans are changing the weather') are intuitively dubious, yet reputable scientists have convinced many laypeople of their truth.

In this study, we intentionally selected authorities that are generally considered benevolent[30,31] and we generated statements that are nearly impossible to (in)validate and that bear no relation to controversial or politicized scientific topics about which people may have strong prior attitudes (such as efficacy of vaccinations, climate change etc.). By using ambiguous claims without any specific ideological content, we tried to isolate the worldview effect regarding the source from any worldview effect related to the content of the claims. At the same time, we aimed to maximize the efficacy of our manipulation, by varying the names, photographs, and visual contexts (chalkboard vs. stars) in addition to the authorities' profession. This approach makes it more difficult to single out which specific factor contributes to the source effect (e.g., the observed effects might be partly driven by the authorities' appearance rather than their domain of expertise). Relatedly, some participants might have recognized the depicted men (Enrico Fermi and José Argüelles), although we consider it unlikely that many did. As we did not ask whether participants recognized any of the depicted sources, we tried to indirectly and retrospectively assess recognition by scanning the open text items at the end of the survey (comments and awareness item) for any mentioning of either 'Enrico', 'Fermi', 'José', or 'Argüelles' (ignoring capitalization or diacritical marks). Only one (Spanish) participant mentioned recognizing both of the sources. While this obviously does not prove no other participants might have known the depicted sources, it seems unlikely that this was the case for a large proportion of participants. On the other hand, the multifaceted nature of the manipulation also increases its ecological validity; our stimuli resemble popular internet memes and real-life instances of source credibility also involve a combination of different features (e.g., authorities typically look the part in public and appear in congruous contexts). Furthermore, a recent study showed that the mere mentioning of a famous source such as Aristotle or the Dalai Lama enhanced profundity ratings for pseudo-profound nonsense relative to unauthored versions, suggesting that even the mere name of an authority may suffice to induce source effects[87].

The effects observed in our experimental data and the associations identified in the existing trust data were highly comparable, suggesting that by using our source credibility manipulation we tapped into participants' attitudes about scientific and religious authorities. A noteworthy divergence, however, is that whereas our data showed a small positive relation between religiosity and credibility ratings for gobbledegook from the scientist, the trust data demonstrated a small but negative association between religiosity and trust in scientists. The finding that religious people are generally less trusting towards science has often been reported in the literature[53,88–90]. However, recent studies suggest that the negative relation between religiosity and trust in science might be US-specific and be weak or absent in other countries[91–94]. Additionally, although trust is likely closely linked to credibility, explicit trust assessments and credibility ratings of specific statements may diverge, perhaps particularly for the kind of obscure statements used in the current study. That is, the gobbledegook statements may still have resonated better with religious individuals than non-religious individuals, resulting in the main effect of religiosity on credibility ratings. This main effect may be driven by a tendency for intuitive reasoning, which has been related to religiosity[78,95,96] and receptivity of pseudo-profound and pseudo-scientific nonsense[36,67]. It could thus be that mistrust in science only partially dampens the allure of well-sounding science-related gobbledegook for intuitive reasoners[36].

Notably, our study showed that across 24 countries even those who are highly religious are prone to a scientific source credibility bias, what we have deemed the Einstein effect. Looking ahead, there are at least six compelling horizons for future research to address the generalizability and underlying causes of the Einstein effect. First, whether scientific education diminishes the appeal of scientific authority outside its immediate domain remains unclear. Although those who place faith in science are prone to Einstein effects[38,40,97,98], strong scepticism is normative within the practice of science – as anyone who has experienced peer-review will attest. Although it is 150 years after Charles Peirce famously argued for fixing beliefs from the "method of science" in favour the "method of authority" the role of appeals to scientific authority among scientists remains unclear[99]. Second, future researchers might investigate whether political partisanship predicts differences in scientific-source credibility. Although political commitments may share common psychological features with religious commitments[100–103], the rise of anti-science populist ideologies might diminish or reverse Einstein effects among political partisans. In contrast, individual differences in deference to science[104] may predict enhanced Einstein effects, although a recent study failed to find this pattern for faith in science[105]. Third, the historical origins of scientific source credibility across different cultures remain unclear. If we were to wind back the clock a century to Einstein's era, would we also observe preferential source-credibility for scientific authority over spiritual authority? Fourth, the proximate and sustaining social and technological causes of scientific source credibility are not addressed in our study, and remain ripe for investigations. Is scientific source credibility an artefact of global information networks, country-wide science education, or the sequestering of religious authority to the private domain? Fifth, although our study covers 24 countries worldwide, we cannot claim universality for our findings. Indeed, investigating source credibility in cultures where spiritual authority dominates may help to clarify the mechanistic questions that our study raises but does not address. Sixth, future work may extend the current work and investigate how the Einstein effect is affected by content cues (e.g., the use of jargon, argument coherence, disclosure of uncertainty[106]) and personal attitudes towards the topic[107–109].

In conclusion, our results strongly suggest that scientific authority is generally considered a reliable source for truth, more so than spiritual authority. Indeed, there are ample examples demonstrating that science serves as an important cue for credibility; the cover of Donald Trump's niece's family history book is adorned by "Mary L. Trump, PhD"; advertisements for cosmetic products often claim to be "clinically proven" and "recommended by dermatologists", and even the tobacco industry used to appeal to science (e.g., "more doctors smoke Camels than any other cigarette"). By systematically quantifying the difference between acceptance of statements by a scientific and spiritual authority in a global sample, this work addresses the fundamental question of how people trust what others say about the world. Although science and scientists are certainly not infallible, it may be reassuring that irrespective of one's religious worldview, most people still have a tendency to use science more than spirituality as a heuristic for the trustworthiness and credibility of information.

Methods

Participants

In total, $10,535$ participants completed the online experiment. Of these, 340 participants (3.23%) were excluded because they failed the attention check (but see Table 2 for equivalent results when data all participants are included), leaving an analytic sample of $N = 10,195$ from 24 countries (see Table 3 for descriptive statistics per country). Participants were recruited from university student samples, from personal networks, and from representative samples accessed by panel agencies and online platforms (MTurk, Kieskompas, Sojump, TurkPrime, Lancers, Qualtrics panels, Crowdpanel, and Prolific). Participants were compensated for participation by a financial remuneration, the possibility for a reward through a raffle, course credits, or no compensation. There were no a priori exclusion criteria; everyone over 18 years old could participate. Participants were forced to answer all multiple choice questions, hence there was no missing data (except for 36 people who did not provide a valid age). The countries were convenience-sampled (i.e., through personal networks), but were selected to cover all 6 continents and include different ethnic majorities and religious majorities (Christian, Muslim, Hindu, Jewish, Eastern religions, as well as highly secular societies). Table 3 displays the method of recruitment and compensation per country.

The study was approved by the local ethics committee at the Psychology Department of the University of Amsterdam (Project #2018-SP-9713). Additional approval was obtained from local IRBs at the Adolfo Ibáñez University (Chile), the Babes-Bolyai University (Romania), the James Cook University (Singapore), Royal Holloway, University of London (UK), and the University of Connecticut (US).

Table 3: Descriptives Statistics per Country

	N	Age (SD)	Women (%)	Religiosity	Sample	Compensation
Australia	463	48.3 (16.0)	48.4	0.52	online panel	money
Belgium	320	34.6 (13.1)	55.6	0.24	mixed	raffle
Brazil	402	28.8 (10.4)	73.1	0.51	mixed	none; credits
Canada	351	33.2 (10.5)	52.4	0.28	online panel	money
Chile	308	30.8 (9.9)	59.1	0.33	mixed	raffle
China	390	32.1 (8.4)	55.9	0.32	online panel	money
Croatia	309	28.0 (6.9)	78.3	0.41	mixed	raffle
Denmark	415	27.9 (10.3)	71.3	0.26	mixed	raffle
France	405	40.6 (12.8)	64.2	0.29	online panel	money
Germany	1,287	27.5 (9.0)	62.2	0.32	mixed	raffle
India	394	30.4 (6.5)	36.3	0.73	online panel	money
Ireland	434	42.6 (15.0)	51.8	0.48	online panel	money
Israel	501	27.9 (10.1)	73.5	0.37	students	credits
Italy	342	27.2 (8.2)	50.9	0.26	mixed	none; money
Japan	424	40.6 (10.0)	43.9	0.29	online panel	money
Lithuania	291	24.1 (7.0)	83.2	0.35	students	none
Morocco	329	32.1 (11.8)	16.1	0.70	online panel	money
Netherlands	482	57.6 (14.7)	25.3	0.28	online panel	money
Romania	539	24.4 (7.4)	85.2	0.55	mixed	raffle
Singapore	308	22.2 (3.4)	62.0	0.45	students	credits
Spain	337	41.9 (13.9)	31.2	0.21	online panel	money
Turkey	362	39.2 (11.1)	24.6	0.33	online panel	money
UK	400	36.2 (12.7)	65.8	0.23	online panel	money
US	402	35.8 (14.4)	51.0	0.45	mixed	none; money
Total	10,195	33.8 (13.8)	55.9	0.38	–	–

Note. Religiosity refers to the self-reported level of individual religiosity, transformed on a 0-1 scale. Sample indicates the composition of the sample based on the method of recruitment per site.

Sampling Plan

We preregistered a target sample size of $n = 400$ per country and 20-25 target countries. The preregistered sample size and composition allowed us to look at overall effects, effects within countries, and between countries. As we applied a Bayesian statistical framework, we needed a minimum of 20 countries to have sufficient data for accurate estimation in cross-country comparisons[110]. However, our main interest were overall effects - rather than effects for individual countries. With approximately 8,800 participants, we would have sufficient data to reliably estimate overall effects, especially since the source effect is within-subjects. Data collection was terminated by November 30th, 2019. The data from ten participants who completed the survey after this termination date were retained in the dataset.

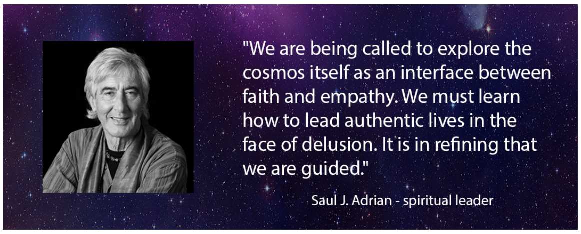
Materials

The study was part of a larger project on cross-cultural effects related to religiosity (see Appendix for details about the project). The full translated survey for each included country can be found at osf.io/kywjs/. The relevant variables for the current study were individual religiosity, the manipulated source of authority, and the ratings of the statements.

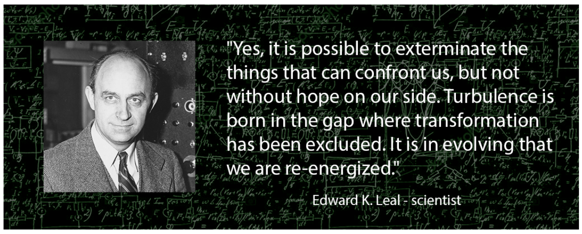
Participant religiosity was measured using established items taken from the World Values Survey[80], covering religious behaviours (institutionalized such as church attendance and private such as prayer/mediation), beliefs, identification, values, and denomination (see Table A5 for the exact items). Besides having high face-validity, these measures have been applied cross-culturally in other studies[79,111,112]. A Bayesian reliability analysis using the `Bayesrel` package[113] indicated good internal consistency of the religiosity measure, McDonald omega = 0.930 $[0.927, 0.931]$. The religious membership item was removed from the scale, as this item was only moderately correlated with the other items (item-rest correlation = 0.608, all others > 0.706) and dropping it improved the reliability to omega = 0.939 $[0.938, 0.941]$. The remaining seven individual religiosity items were transformed on a 0-1 scale (to make each item contribute equally to the scale), tallied to create a religiosity score per participant, and grand-mean standardized for the analyses.

The experimental stimuli consisted of two gobbledegook statements that were attributed to a spiritual guru and to a scientific authority (within-subjects). We created two versions of the statement, manipulating (1) the background of the frame: an opaque new-age purple galaxy background vs. an opaque dark green chalkboard with physics equations, (2) the accompanying gray-scale photo of the alleged source: a man in robes (photo of José Argüelles) vs. a man in an old-fashioned suit (photo of Enrico Fermi), and (3) the reported profession: spiritual leader vs. scientist (see Figure 5)). Additionally, in the introductory text, the source was further announced as "Saul J. Adrian - a spiritual authority in world religions" vs. "Edward K. Leal - a scientific authority in the field of particle physics", names counter-balanced. The names were fictitious and the photos were taken from Wikipedia with re-use permission. The two versions of the text were three-sentence, 37/38 word statements. We generated the statements using the New-Age bullshit generator (http://sebpearce.com/bullshit/), that combines new-age buzzwords in a syntactically correct structure resulting in meaningless, but pseudo-profound sounding texts[67]. The two versions of the text were counterbalanced between sources. Participants were randomly assigned to the scientific-spiritual or the spiritual-scientific ordered condition. The stimuli in each language are provided at osf.io/qsyvw/.

The main outcome variable pertained to judgments of importance and credibility of gobbledegook, measured on a 7-point Likert scale from *not at all important / not at all credible* to *extremely important / extremely credible*, respectively. A multiple choice recognition item for the source that expressed the statement was included as a manipulation check. In our preregistration, we did not specify that we would exclude participants based on incorrect recall of the source of the statement. We therefore kept all observations in the data set for the main analyses and additionally ran the models without the observations for which the source was not recalled correctly. The results of this robustness check are provided in Table 2. For exploratory purposes, we also measured reading and processing time for the statement, as well as depth of processing.



(a) Spiritual authority – Statement A



(b) Scientific authority – Statement B

Figure 5: Example stimuli used in the survey. The statements were generated using the New-Age bullshit generator (http://sebpearce.com/bullshit/) and translated into the language the study was conducted in. The statements were counterbalanced between sources across participants.

The latter was operationalized as the number of items correctly identified as having appeared in the statement. Participants were presented with a list of 10 words, including 5 targets and 5 distractors, and were asked to select the words that they recognised from the statement.

Procedure

Participants received a link to the Qualtrics survey, either by email, social media or through an online platform. After reading the instructions and providing informed consent, they first completed items for a separate study about religiosity and trustworthiness. Next, they were presented with the first statement and source stimulus, rated its importance and credibility, completed the manipulation check to validate that they registered the source, and completed the word recall item. These elements were then repeated for the second statement. After that, participants completed items about body-mind dualism. Finally, they provided demographics, a quality of life scale, the religiosity items and were given the opportunity to provide comments. It took about 10 minutes to complete the entire survey (median completion time was 11.4 minutes).

Data Analysis

We used the R package `BayesFactor`[76] to estimate and test the multilevel Bayesian regression models[114,115]. The multilevel Bayesian modelling approach allows us to systematically evaluate the evidence in the data under different models: (i) across all countries the effect is truly null; (ii) all countries share a common nonzero effect; (iii) countries differ, but all effects are in the same (predicted) direction; and (iv) in some countries the effect is positive whereas in others the effect is negative. The models differ in the extent to which they constrain their predictions, from the most constrained (i) to completely unconstrained (iv). We refer to these models as the null model, the common effect model, the positive effects model, and the unconstrained model, respectively. Note that while the predictions from model (iii) are less constrained than those from model (ii), it is more difficult to obtain evidence for small effects under the latter model because it assumes that the effect is present in every country, rather than only in the aggregate sample. When applied to our hypothesis for the source effect, evidence for (i) would indicate that people from these 24 countries do not differentially evaluate credibility of claims from a guru or a scientist, evidence for (ii) would indicate that on average people from these 24 countries consider claims from a scientist more credible than from a guru (or vice versa) with little between-country variability in the size of the effect, evidence for (iii) would indicate that in all of the 24 countries, people consider claims from a scientist more credible than from a guru (or vice versa), but there is cultural variation in the size of this effect, and evidence for (iv) would indicate that in some countries people consider claims from a scientist more credible than from a guru, and in other countries people consider claims from a guru more credible than from a scientist, indicating cultural variation in the direction (and size) of the effect. We used the interpretation categories for Bayes factors proposed by Lee and Wagenmakers[116], based on the original labels specified by Jeffreys[117].

For the main effect of source ($\mathcal{H}_1$), we specified the following unconstrained model. Let Y_{ijk} be the credibility rating for the ith participant, $i = 1, ..., N$, in the jth country, $j = 1, ..., 24$, for the kth condition, $k = 1, 2$. Then:

$$Y_{ijk} \sim N(\mu + \alpha_j + v_i\beta + r_i\delta_j + x_k\gamma_j, \sigma^2).$$

Here, the term $\mu + \alpha_j$ serve as the baseline credibility intercepts with μ being the grand mean and α_j the jth country's deviation from the grand mean. The β term reflects the fixed effect of the level of education covariate. δ_j is the jth country's main effect of religiosity on credibility ratings. The crucial parameter here is γ_j which is the source effect for the jth country. In the common effects model, we will replace γ_i with γ. The variable $x_k = -0.5, 0.5$ if $k = 1, 2$, respectively, where $k = 1$ indicates the scientist condition and the $k = 2$ indicates the guru condition. The variable v_i is the standardized participant-level education covariate. The variable r_i is the standardized religiosity score for each participant. Finally, σ^2 is the variance in credibility ratings across participants.

To test the source-by-religiosity interaction for hypothesis 2, the model from (1) is extended by including an interaction term:

$$Y_{ijk} \sim N(\mu + \alpha_j + v_i\beta + r_i\delta_j + x_k\gamma_j + r_i x_k\theta_j, \sigma^2),$$

where θ_j is the parameter of interest, the religiosity*source interaction effect, with $r_i x_k$ as the product of the experimental condition and the standardized individual religiosity score. The parameter estimates as reported in the results section are based on the full model from (2).

In order to systematically investigate which third variables should and should not be included in the statistical model, we used *directed acyclic graphs* (DAGs[118]) to visually represent the causal relations between the variables in our data[119–121]. In short, this method entails specifying directed relations (arrows) between different constructs and measures (nodes) in a given design, that allow one to intuitively reflect causal structures and determine which third variables should be accounted for and which should be ignored in the statistical model. Based on DAGs created in the R package `ggdag`[122], both *country* and *level of education* were identified as potential confounding factors that warranted inclusion, as they may affect both religiosity[123,124] and overall credibility assessments (e.g., due to skepticism). Country was therefore added as a clustering factor, while level of education was added as a fixed covariate in all models. We also ran the models while including all participant-level variables related to the primary measures, i.e., gender[125], age[126], SES[127,128], statement version (A or B), and presentation order (guru–scientist or scientist–guru). Note that including these covariates improved the model fit, but the qualitative results remain the same regardless of the (set of) covariates. See Figures A4-A6 for details on the causal graphs and Table 2 for the primary results without any and with all covariates.

Prior Settings

The BayesFactor package applies the default priors for ANOVA and regression designs[129,130], in which the researcher can determine the scale settings for each individual predictor in the model. We used the settings for the critical priors in the multilevel models as proposed by Rouder et al.[115], concerning the scale settings on μ_γ, μ_θ and $\sigma^2_\gamma, \sigma^2_\theta$. The scale on μ_γ, μ_θ reflects the expected size of the overall source effect and source-by-religiosity effect, respectively, and is set to 0.4 (small-medium effect). The scale of $\sigma^2_\gamma, \sigma^2_\theta$ reflects the expected amount of variability in these effects across countries. This scale is set to 60% of the overall effect, resulting in a value of 0.24. The prior scale for the overall between-countries variance was set to 1. We used 31,000 iterations for the Markov chain Monte Carlo sampling and discarded the first 1,000 iterations ("burn-in").

Deviations from Preregistration

We deviated from the preregistration in the following ways. First, in our preregistration, we formulated a hypothesis about the interaction between source and perceived cultural norms of religiosity in one's country. However, in retrospect, we realized this hypothesis lacked theoretical justification and the proposed analysis was methodologically suboptimal (see Appendix for details on this analysis).

Second, as a stopping rule, we preregistered that data collection would be terminated (a) when the target of $n = 400$ per country was reached, or (b) by September 30th, 2019. However, due to unforeseen delays in construction of the materials and recruitment, this deadline was extended until November 30th, 2019. We did not download or inspect the data until after November 30th.

Third, we preregistered to only include countries where usable data from at least 300 participants were collected (i.e., complete data from attentive participants). However, we decided to keep the $n = 291$ participants from Lithuania in the final sample, as the hierarchical models account for uncertainty in estimates from countries with smaller samples and removing these data will actually reduce the overall precision of the estimates. Moreover, it would simply be unfortunate to remove all data from a highly understudied country.

Fourth, we preregistered that we would use the R package brms[131] to analyse the data and estimate model parameters. However, we ended up using the BayesFactor package[76]. This method is arguably more suitable for model comparison and calculating Bayes factors in particular. However, we also ran the models as preregistered and report these results in the Appendix.

Fifth, we added level of education as a participant-level covariate to the models, which improved the model fits. Note that adjustments 3-5 did not qualitatively change any of the results (see Table 2 and the Appendix).

References

1. The White House Press Briefing. *Remarks by President Trump in Press Briefing [Statements & Releases]* https://www.whitehouse.gov/briefings-statements/remarks-president-trump-press-briefing-092420/. Press Release. 2020.
2. Farr, C. *Stanford Medical Faculty Lambaste Former Colleague and Trump Coronavirus Advisor Dr. Scott Atlas* https://www.cnbc.com/2020/09/10/stanford-medical-faculty-lambaste-former-colleague-and-trump-coronavirus-advisor-dr-scott-atlas.html. 2020.
3. Brinol, P. & Petty, R. E. Source Factors in Persuasion: A Self-Validation Approach. *European Review of Social Psychology* **20,** 49–96. doi:10.1080/10463280802643640 (2009).
4. Pornpitakpan, C. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of applied social psychology* **34,** 243–281. doi:10.1111/j.1559-1816.2004.tb02547.x (2004).
5. McGinnies, E. & Ward, C. D. Better Liked than Right: Trustworthiness and Expertise as Factors in Credibility. *Personality and Social Psychology Bulletin* **6,** 467–472. doi:10.1177/014616728063023 (1980).
6. Harris, A. J., Hahn, U., Madsen, J. K. & Hsu, A. S. The Appeal to Expert Opinion: Quantitative Support for a Bayesian Network Approach. *Cognitive Science* **40,** 1496–1533 (2016).
7. Smith, C. T., De Houwer, J. & Nosek, B. A. Consider the Source: Persuasion of Implicit Evaluations Is Moderated by Source Credibility. *Personality and Social Psychology Bulletin* **39,** 193–205. doi:10.1177/0146167212472374 (2013).
8. Sperber, D. *et al.* Epistemic Viligance. *Mind & Language* **25,** 359–393. doi:10.1111/j.1468-0017.2010.01394.x (2010).
9. Chaiken, S. & Maheswaran, D. Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment. *Journal of Personality and Social Psychology* **66,** 460–473. doi:10.1037/0022-3514.66.3.460 (1994).
10. Petty, R. E. & Cacioppo, J. T. in *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* 1–24 (Springer-Verlag, New York, NY, 1986). doi:10.1007/978-1-4612-4964-1_1.
11. Clément, F., Koenig, M. & Harris, P. The Ontogenesis of Trust. *Mind & Language* **19,** 360–379. doi:10.1111/j.0268-1064.2004.00263.x (2004).
12. Jaswal, V. K. & Neely, L. A. Adults Don't Always Know Best: Preschoolers Use Past Reliability over Age When Learning New Words. *Psychological Science* **17,** 757–758. doi:10.1111/j.1467-9280.2006.01778.x (2006).
13. Taylor, M., Cartwright, B. S. & Bowden, T. Perspective Taking and Theory of Mind: Do Children Predict Interpretive Diversity as a Function of Differences in Observers' Knowledge? *Child Development* **62,** 1334–1351. doi:10.1111/j.1467-8624.1991.tb01609.x (1991).

14. Harris, P. L., Koenig, M. A., Corriveau, K. H. & Jaswal, V. K. Cognitive Foundations of Learning from Testimony. *Annual Review of Psychology* **69,** 251–273. doi:10.1146/annurev-psych-122216-011710 (2018).

15. Birch, S. A. J., Akmal, N. & Frampton, K. L. Two-Year-Olds Are Vigilant of Others' Non-Verbal Cues to Credibility. *Developmental Science* **13,** 363–369. doi:10.1111/j.1467-7687.2009.00906.x (2010).

16. Chudek, M., Heller, S., Birch, S. & Henrich, J. Prestige-Biased Cultural Learning: Bystander's Differential Attention to Potential Models Influences Children's Learning. *Evolution and Human Behavior* **33,** 46–56. doi:10.1016/j.evolhumbehav.2011.05.005 (2012).

17. Mascaro, O. & Sperber, D. The Moral, Epistemic, and Mindreading Components of Children's Vigilance towards Deception. *Cognition* **112,** 367–380. doi:10.1016/j.cognition.2009.05.012 (2009).

18. Johnston, A. M., Mills, C. M. & Landrum, A. R. How Do Children Weigh Competence and Benevolence When Deciding Whom to Trust? *Cognition* **144,** 76–90. doi:10.1016/j.cognition.2015.07.015 (2015).

19. Fiske, S. T. & Dupree, C. Gaining Trust as Well as Respect in Communicating to Motivated Audiences about Science Topics. *Proceedings of the National Academy of Sciences* **111,** 13593–13597. doi:10.1073/pnas.1317505111 (2014).

20. Bloom, P. & Weisberg, D. S. Childhood Origins of Adult Resistance to Science. *Science* **316,** 996–997. doi:10.1126/science.1133398 (2007).

21. Harris, P. L. *Trusting What You're Told: How Children Learn from Others* (Harvard University Press, 2012).

22. Mercier, H. *Not Born Yesterday: The Science of Who We Trust and What We Believe.* (Princeton University Press, 2020).

23. Hahn, U., Harris, A. J. L. & Corner, A. Public Reception of Climate Science: Coherence, Reliability, and Independence. *Topics in Cognitive Science* **8,** 180–195 (2016).

24. Henrich, J. & Gil-White, F. J. The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission. *Evolution and Human Behavior* **22,** 165–196. doi:10.1016/s1090-5138(00)00071-4 (2001).

25. Henrich, J. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter.* (Princeton University Press, 2015).

26. Johnson, D. D. & Fowler, J. H. The Evolution of Overconfidence. *Nature* **477,** 317–320 (2011).

27. Johnson, D. D. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics* (Princeton University Press, 2020).

28. Sperber, D. Intuitive and Reflective Beliefs. *Mind & Language* **12,** 67–83. doi:10.1111/j.1468-0017.1997.tb00062.x (1997).

29. Sperber, D. The Guru Effect. *Review of Philosophy and Psychology* **1,** 583–592. doi:10.1007/s13164-010-0025-0 (2010).

30. Funk, C. *Key Findings about Americans' Confidence in Science and Their Views on Scientists' Role in Society* https://pewrsr.ch/2Hgq31S. 2020.

31. Krause, N. M., Brossard, D., Scheufele, D. A., Xenos, M. A. & Franke, K. Trends—Americans' Trust in Science and Scientists. *Public Opinion Quarterly* **83,** 817–836. doi:10.1093/poq/nfz041 (2019).

32. Reynolds, T. W., Bostrom, A., Read, D. & Morgan, M. G. Now What Do People Know about Global Climate Change? Survey Studies of Educated Laypeople. *Risk Analysis: An International Journal* **30,** 1520–1538. doi:10.1111/j.1539-6924.2010.01448.x (2010).

33. Weber, E. U. & Stern, P. C. Public Understanding of Climate Change in the United States. *American Psychologist* **66,** 315–328. doi:10.1037/a0023253 (2011).

34. McCloskey, M., Washburn, A. & Felch, L. Intuitive Physics: The Straight-down Belief and Its Origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **9,** 636–649. doi:10.1037/0278-7393.9.4.636 (1983).

35. Mercier, H. The Argumentative Theory: Predictions and Empirical Evidence. *Trends in Cognitive Sciences* **20,** 689–700. doi:10.1016/j.tics.2016.07.001 (2016).

36. Evans, A. M., Sleegers, W. & Mlakar, Ž. *Individual Differences in Receptivity to Scientific Bullshit* Preprint (PsyArXiv, 2020). doi:10.31234/osf.io/2r65q.

37. Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E. & Gray, J. R. The Seductive Allure of Neuroscience Explanations. *Journal of Cognitive Neuroscience* **20,** 470–477. doi:10.1162/jocn.2008.20040 (2008).

38. Fernandez-Duque, D., Evans, J., Christian, C. & Hodges, S. D. Superfluous Neuroscience Information Makes Explanations of Psychological Phenomena More Appealing. *Journal of Cognitive Neuroscience* **27,** 926–944. doi:10.1162/jocn_a_00750 (2014).

39. McCabe, D. P. & Castel, A. D. Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning. *Cognition* **107,** 343–352. doi:10.1016/j.cognition.2007.07.017 (2008).

40. Eriksson, K. The Nonsense Math Effect. *Judgment and Decision Making* **7,** 746–749 (2012).

41. Gruber, D. & Dickerson, J. A. Persuasive Images in Popular Science: Testing Judgments of Scientific Reasoning and Credibility. *Public Understanding of Science* **21,** 938–948. doi:10.1177/0963662512454072 (2012).

42. Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G. & Garry, M. On the (Non) Persuasive Power of a Brain Image. *Psychonomic Bulletin & Review* **20,** 720–725. doi:10.3758/s13423-013-0391-6 (2013).

43. Legare, C. H., Evans, E. M., Rosengren, K. S. & Harris, P. L. The Coexistence of Natural and Supernatural Explanations across Cultures and Development. *Child Development* **83,** 779–793. doi:10.1111/j.1467-8624.2012.01743.x (2012).

44. Van Leeuwen, N. Religious Credence Is Not Factual Belief. *Cognition* **133,** 698–715. doi:10.1016/j.cognition.2014.08.015 (2014).

45. Johnson, K. A., Okun, M. A., Cohen, A. B., Sharp, C. A. & Hook, J. N. Development and Validation of the Five-Factor LAMBI Measure of God Representations. *Psychology of Religion and Spirituality.* doi:10.1037/rel0000207 (2018).

46. Boyer, P. *Religion Explained: The Evolutionary Origins of Religious Thought* (Basic Books, 2001).

47. Friesen, J. P., Campbell, T. H. & Kay, A. C. The Psychological Advantage of Unfalsifiability: The Appeal of Untestable Religious and Political Ideologies. *Journal of Personality and Social Psychology* **108,** 515–529. doi:10.1037/pspp0000018 (2015).

48. Liquin, E. G., Metz, S. E. & Lombrozo, T. Science Demands Explanation, Religion Tolerates Mystery. *Cognition* **204,** 104398. doi:10.1016/j.cognition.2020.104398 (2020).

49. Mahoney, A. Theological Expressions as Costly Signals of Religious Commitment. *The evolution of religion: Studies, theories, and critiques,* 161–166 (2008).

50. Irons, W. Why People Believe (What Other People See as) Crazy Ideas. *The evolution of religion: Studies, theories, and critiques, ed. J. Bulbulia, R. Sosis, C. Genet, R. Genet & K. Wyman,* 51–57 (2008).

51. Lachapelle, E., Montpetit, É. & Gauvin, J.-P. Public Perceptions of Expert Credibility on Policy Issues: The Role of Expert Framing and Political Worldviews - Expert Framing and Political Worldviews. *Policy Studies Journal* **42,** 674–697. doi:10.1111/psj.12073 (2014).

52. Gauchat, G. The Cultural Authority of Science: Public Trust and Acceptance of Organized Science. *Public Understanding of Science* **20,** 751–770. doi:10.1177/0963662510365246 (2011).

53. Gauchat, G. Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010. *American Sociological Review* **77,** 167–187. doi:10.1177/0003122412438225 (2012).

54. Brandt, M. J. & Crawford, J. T. in *Advances in Experimental Social Psychology* (ed Gawronski, B.) 1–66 (Academic Press, 2020). doi:10.1016/bs.aesp.2019.09.002.

55. Levy, N. Due Deference to Denialism: Explaining Ordinary People's Rejection of Established Scientific Findings. *Synthese* **196,** 313–327. doi:10.1007/s11229-017-1477-x (2019).

56. Schjoedt, U., Stødkilde-Jørgensen, H., Geertz, A. W., Lund, T. E. & Roepstorff, A. The Power of Charisma-Perceived Charisma Inhibits the Frontal Executive Network of Believers in Intercessory Prayer. *Social Cognitive and Affective Neuroscience* **6,** 119–127. doi:10.1093/scan/nsq023 (2011).

57. McPhetres, J. & Zuckerman, M. Religious People Endorse Different Standards of Evidence When Evaluating Religious versus Scientific Claims. *Social Psychological and Personality Science* **8,** 836–842. doi:10.1177/1948550617691098 (2017).

58. Lobato, E. J. C., Tabatabaeian, S., Fleming, M., Sulzmann, S. & Holbrook, C. Religiosity Predicts Evidentiary Standards. *Social Psychological and Personality Science*. doi:10.1177/1948550619869613 (2019).

59. Robinson, C. Cross-Cutting Messages and Political Tolerance: An Experiment Using Evangelical Protestants. *Political Behavior* **32,** 495–515. doi:10.1007/s11109-010-9118-9 (2010).

60. Munro, G. D. & Ditto, P. H. Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information. *Personality and Social Psychology Bulletin* **23,** 636–653. doi:10.1177/0146167297236007 (1997).

61. Tversky, A. & Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **185,** 1124–1131. doi:10.1126/science.185.4157.1124 (1974).

62. Cook, J. & Lewandowsky, S. Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks. *Topics in Cognitive Science* **8,** 160–179. doi:10.1111/tops.12186 (2016).

63. Jern, A., Chang, K.-m. K. & Kemp, C. Belief Polarization Is Not Always Irrational. *Psychological Review* **121,** 206–224. doi:10.1037/a0035941 (2014).

64. Hahn, U., Harris, A. J. L. & Corner, A. Argument Content and Argument Source: An Exploration. *Informal Logic* **29,** 337–367. doi:10.22329/il.v29i4.2903 (2009).

65. Batson, C. D. Rational Processing or Rationalization? The Effect of Disconfirming Information on a Stated Religious Belief. *Journal of Personality and Social Psychology* **32,** 176–184. doi:10.1037/h0076771 (1975).

66. Ibrahim, J. G. & Chen, M.-H. Power Prior Distributions for Regression Models. *Statistical Science* **15,** 46–60 (2000).

67. Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J. & Fugelsang, J. A. On the Reception and Detection of Pseudo-Profound Bullshit. *Judgment and Decision Making* **10,** 549–563 (2015).

68. Zinnbauer, B. J. *et al.* Religion and Spirituality: Unfuzzying the Fuzzy. *Journal for the Scientific Study of Religion* **36,** 549–564. doi:10.2307/1387689 (1997).

69. Paloutzian, R. F. & Park, C. L. *Handbook of the Psychology of Religion and Spirituality.* (Guilford Publications, 2014).

70. Lindeman, M., van Elk, M., Lipsanen, J., Marin, P. & Schjødt, U. Religious Unbelief in Three Western European Countries: Identifying and Characterizing Unbeliever Types Using Latent Class Analysis. *The International Journal for the Psychology of Religion* **29,** 184–203. doi:10.1080/10508619.2019.1591140 (2019).

71. McClintock, C. H., Lau, E. & Miller, L. Phenotypic Dimensions of Spirituality: Implications for Mental Health in China, India, and the United States. *Frontiers in Psychology* **7,** 1600. doi:10.3389/fpsyg.2016.01600 (2016).

72. Wilson, M. S., Bulbulia, J. & Sibley, C. G. Differences and Similarities in Religious and Paranormal Beliefs: A Typology of Distinct Faith Signatures. *Religion, Brain & Behavior* **4,** 104–126. doi:10.1080/2153599x.2013.779934 (2013).

73. Rice, T. W. Believe It or Not: Religious and Other Paranormal Beliefs in the United States. *Journal for the Scientific Study of Religion* **42,** 95–106. doi:10.1111/1468-5906.00163 (2003).

74. Gallup. *Wellcome Gobal Monitor – First Wave Findings* https://wellcome.ac.uk/reports/wellcome-global-monitor/2018. 2019.

75. van Doorn, J., Ly, A., Marsman, M. & Wagenmakers, E.-J. Bayesian Rank-Based Hypothesis Testing for the Rank Sum Test, the Signed Rank Test, and Spearman's ρ. *Journal of Applied Statistics* **47,** 2984–3006. doi:10.1080/02664763.2019.1709053 (2020).

76. Morey, R. D. & Rouder, J. N. *BayesFactor: Computation of Bayes Factors for Common Designs* 2018.

77. Mummolo, J. & Peterson, E. Demand Effects in Survey Experiments: An Empirical Assessment. *American Political Science Review* **113,** 517–529. doi:10.1017/S0003055418000837 (2019).

78. Gervais, W. M. *et al.* Analytic Atheism: A Cross-Culturally Weak and Fickle Phenomenon? *Judgment and Decision Making* **13,** 268–274 (2018).

79. Stavrova, O. Religion, Self-Rated Health, and Mortality: Whether Religiosity Delays Death Depends on the Cultural Context. *Social Psychological and Personality Science* **6,** 911–922. doi:10.1177/1948550615593149 (2015).

80. World Values Survey. *Wave 6 Official Aggregate v. 20150418* worldvaluessurvey.org. 2010.

81. Mitkidis, P., Xygalatas, D., Buttrick, N., Porubanova, M. & Lienard, P. The Impact of Authority on Cooperation: A Cross-Cultural Examination of Systemic Trust. *Adaptive Human Behavior and Physiology* **1,** 341–357 (2015).

82. Inglehart, R. Mapping Global Values. *Comparative Sociology* **5,** 115–136. doi:10.1163/156913306778667401 (2006).

83. Singelis, T. M., Triandis, H. C., Bhawuk, D. P. S. & Gelfand, M. J. Horizontal and Vertical Dimensions of Individualism and Collectivism: A Theoretical and Measurement Refinement. *Cross-Cultural Research* **29,** 240–275. doi:10.1177/106939719502900302 (1995).

84. Sturgis, P., Roberts, C. & Smith, P. Middle Alternatives Revisited: How the Neither/nor Response Acts as a Way of Saying "I Don't Know"? *Sociological Methods & Research* **43,** 15–38. doi:10.1177/0049124112452527 (2014).

85. Raaijmakers, Q. A., Van Hoof, J., t Hart, H., Verbogt, T. & Vollebergh, W. A. Adolescents' Midpoint Responses on Likert-Type Scale Items: Neutral or Missing Values? *International Journal of Public Opinion Research* **12,** 208–216 (2000).

86. Krosnick, J. A. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* **5,** 213–236. doi:10.1002/acp.2350050305 (1991).

87. Gligorić, V. & Vilotijević, A. "Who Said It?" How Contextual Information Influences Perceived Profundity of Meaningful Quotes and Pseudo-Profound Bullshit. *Applied Cognitive Psychology* **34,** 535–542. doi:10.1002/acp.3626 (2020).

88. O'Brien, T. L. & Noy, S. Cultural Authority in Comparative Context: A Multilevel Analysis of Trust in Science and Religion. *Journal for the Scientific Study of Religion* **57,** 495–513. doi:10.1111/jssr.12537 (2018).

89. McPhetres, J. & Zuckerman, M. Religiosity Predicts Negative Attitudes towards Science and Lower Levels of Science Literacy. *PLoS One* **13,** e0207125. doi:10.1371/journal.pone.0207125 (2018).

90. Farias, M., Newheiser, A.-K., Kahane, G. & de Toledo, Z. Scientific Faith: Belief in Science Increases in the Face of Stress and Existential Anxiety. *Journal of Experimental Social Psychology* **49,** 1210–1213. doi:10.1016/j.jesp.2013.05.008 (2013).

91. Cacciatore, M. A. *et al.* Opposing Ends of the Spectrum: Exploring Trust in Scientific and Religious Authorities. *Public Understanding of Science* **27,** 11–28. doi:10.1177/0963662516661090 (2018).

92. McPhetres, J., Jong, J. & Zuckerman, M. Religious Americans Have Less Positive Attitudes toward Science, but This Does Not Extend to Other Cultures. *Social Psychological and Personality Science,* 1–9. doi:10.1177/1948550620923239 (2020).

93. Rutjens, B. T. & van der Lee, R. Spiritual Skepticism? Heterogeneous Science Skepticism in the Netherlands. *Public Understanding of Science* **29,** 335–352. doi:10.1177/0963662520908534 (2020).

94. Rutjens, B. T. *et al.* Science Skepticism Across 24 Countries. *Social Psychological and Personality Science,* 19485506211001329. doi:10.1177/19485506211001329 (2021).

95. Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J. & Fugelsang, J. A. Analytic Cognitive Style Predicts Religious and Paranormal Belief. *Cognition* **123,** 335–346. doi:10.1016/j.cognition.2012.03.003 (2012).

96. Pennycook, G., Ross, R. M., Koehler, D. J. & Fugelsang, J. A. Atheists and Agnostics Are More Reflective than Religious Believers: Four Empirical Studies and a Meta-Analysis. *PLOS ONE* **11,** e0153039. doi:10.1371/journal.pone.0153039 (2016).

97. Macdonald, K., Germine, L., Anderson, A., Christodoulou, J. & McGrath, L. M. Dispelling the Myth: Training in Education or Neuroscience Decreases but Does Not Eliminate Beliefs in Neuromyths. *Frontiers in Psychology* **8.** doi:10.3389/fpsyg.2017.01314 (2017).

98. Mayo, R. The Skeptical (Ungullible) Mindset. *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs* **140** (2019).

99. Peirce, C. S. P. in *The Essential Peirce, Volume 1: Selected Philosophical Writings (1867–1893)* (Indiana University Press, 1992).

100. Malka, A., Lelkes, Y., Srivastava, S., Cohen, A. B. & Miller, D. T. The Association of Religiosity and Political Conservatism: The Role of Political Engagement. *Political Psychology* **33,** 275–299. doi:10.1111/j.1467-9221.2012.00875.x (2012).

101. Graham, J., Haidt, J. & Nosek, B. A. Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology* **96,** 1029–1046. doi:10.1037/a0015141 (2009).

102. Johnson, M. K. *et al.* A Mediational Analysis of the Role of Right-Wing Authoritarianism and Religious Fundamentalism in the Religiosity–Prejudice Link. *Personality and Individual Differences* **50,** 851–856. doi:10.1016/j.paid.2011.01.010 (2011).

103. Grafman, J., Cristofori, I., Zhong, W. & Bulbulia, J. The Neural Basis of Religious Cognition. *Current Directions in Psychological Science* **29,** 126–133. doi:10.1177/0963721419898183 (2020).

104. Howell, E. L., Wirz, C. D., Scheufele, D. A., Brossard, D. & Xenos, M. A. Deference and Decision-Making in Science and Society: How Deference to Scientific Authority Goes beyond Confidence in Science and Scientists to Become Authoritarianism. *Public Understanding of Science* **29,** 800–818. doi:10.1177/0963662520962741 (2020).

105. van der Miesen, M. M., van der Lande, G. J., Hoogeveen, S., Schjoedt, U. & van Elk, M. *The Effect of Source Credibility on the Evaluation of Statements in a Spiritual and Scientific Context: A Registered Report Study* in preparation.

106. Corner, A. & Hahn, U. Evaluating Science Arguments: Evidence, Uncertainty, and Argument Strength. *Journal of Experimental Psychology: Applied* **15,** 199–212. doi:10.1037/a0016533 (2009).

107. Scurich, N. & Shniderman, A. The Selective Allure of Neuroscientific Explanations. *PLoS One* **9,** e107529. doi:10.1371/journal.pone.0107529 (2014).

108. Kruglanski, A. W. *et al.* in *Advances in Experimental Social Psychology* 345–392 (San Diego, 2005).

109. Kahan, D. M., Jenkins-Smith, H. & Braman, D. Cultural Cognition of Scientific Consensus. *Journal of Risk Research* **14,** 147–174. doi:10.1080/13669877.2010.511246 (2011).

110. Hox, J. J. C. M., van de Schoot, R. & Matthijsse, S. How Few Countries Will Do? Comparative Survey Analysis from a Bayesian Perspective. *Survey Research Methods* **6,** 87–93. doi:10.18148/srm/2012.v6i2.5033 (2012).

111. Lindeman, M., Svedholm-Hakkinen, A. M. & Lipsanen, J. Ontological Confusions but Not Mentalizing Abilities Predict Religious Belief, Paranormal Belief, and Belief in Supernatural Purpose. *Cognition* **134,** 63–76. doi:10.1016/j.cognition.2014.09.008 (2015).

112. Lun, V. M.-C. & Bond, M. H. Examining the Relation of Religion and Spirituality to Subjective Well-Being across National Cultures. *Psychology of Religion and Spirituality* **5,** 304–315. doi:10.1037/a0033641 (2013).

113. Pfadt, J. M. & van den Bergh, D. *Bayesrel: Bayesian Reliability Estimation* 2020.

114. Haaf, J. M. & Rouder, J. N. Developing Constraint in Bayesian Mixed Models. *Psychological Methods* **22,** 779–798. doi:10.31234/osf.io/ktjnq (2017).

115. Rouder, J. N., Haaf, J. M., Davis-Stober, C. P. & Hilgard, J. Beyond Overall Effects: A Bayesian Approach to Finding Constraints in Meta-Analysis. *Psychological Methods* **24,** 606–621. doi:10.1037/met0000216 (2019).

116. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge (UK), 2013).

117. Jeffreys, H. *Theory of Probability* First (Oxford University Press, Oxford, UK, 1939).

118. Pearl, J. Causal Diagrams for Empirical Research. *Biometrika* **82,** 669–688. doi:10.1093/biomet/82.4.669 (1995).

119. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* doi:10.1201/9781315372495 (Chapman & Hall/CRC Press, Boca Raton (FL), 2016).

120. Pearl, J. The Seven Tools of Causal Inference, with Reflections on Machine Learning. *Communications of the ACM* **62,** 54–60. doi:10.1145/3241036 (2019).

121. Rohrer, J. M. Thinking Clearly about Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* **1,** 27–42. doi:10.1177/2515245917745629 (2018).

122. Barrett, M. *Ggdag: Analyze and Create Elegant Directed Acyclic Graphs* 2021.

123. Albrecht, S. L. & Heaton, T. B. Secularization, Higher Education, and Religiosity. *Review of Religious Research* **26,** 43–58. doi:10.2307/3511041 (1984).

124. Schwadel, P. Does Higher Education Cause Religious Decline? A Longitudinal Analysis of the within- and between-Person Effects of Higher Education on Religiosity. *The Sociological Quarterly* **57,** 759–786. doi:10.1111/tsq.12153 (2016).

125. Miller, A. S. & Hoffmann, J. P. Risk and Religion: An Explanation of Gender Differences in Religiosity. *Journal for the Scientific Study of Religion* **34,** 63–75. doi:10.2307/1386523 (1995).

126. Argue, A., Johnson, D. R. & White, L. K. Age and Religiosity: Evidence from a Three-Wave Panel Analysis. *Journal for the Scientific Study of Religion,* 423–435. doi:10.2307/1387762 (1999).

127. Pyle, R. E. Trends in Religious Stratification: Have Religious Group Socioeconomic Distinctions Declined in Recent Decades? *Sociology of Religion* **67,** 61–79. doi:10.1093/socrel/67.1.61 (2006).

128. Smith, C. & Faris, R. Socioeconomic Inequality in the American Religious System: An Update and Assessment. *Journal for the Scientific Study of Religion* **44,** 95–104. doi:10.1111/j.1468-5906.2005.00267.x (2005).

129. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes Factors for ANOVA Designs. *Journal of Mathematical Psychology* **56,** 356–374. doi:10.1016/J.JMP.2012.08.001 (2012).

130. Rouder, J. N. & Morey, R. D. Default Bayes Factors for Model Selection in Regression. *Multivariate Behavioral Research* **47,** 877–903. doi:10.1080/00273171.2012.734737 (2012).

131. Bürkner, P.-C. Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80,** 1–28. doi:10.18637/jss.v080.i01 (2017).

132. Wan, C. *et al.* Perceived Cultural Importance and Actual Self-Importance of Values in Cultural Identification. *Journal of Personality and Social Psychology* **92,** 337–354. doi:10.1037/0022-3514.92.2.337 (2007).

133. Gabry, J., Goodrich, B. & Lysy, M. *Rstantools: Tools for Developing r Packages Interfacing with Stan.* 2020.

134. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. R-Squared for Bayesian Regression Models. *The American Statistician* **73,** 307–309. doi:10.1080/00031305.2018.1549100 (2019).

135. Gelman, A. *et al. Bayesian Data Analysis (3rd Ed.)* (Chapman & Hall/CRC, Boca Raton (FL), 2014).

136. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved Rˆ for Assessing Convergence of MCMC. *Bayesian Analysis,* 1–28. doi:10.1214/20-BA1221 (2021).

137. Geyer, C. J. in *Handbook of Markov Chain Monte Carlo* (eds Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L.) (CRC Press, 2011). doi:10.1201/b10905-3.

138. Carpenter, B. *We Were Measuring the Speed of Stan Incorrectly—It's Faster than We Thought in Some Cases Due to Antithetical Sampling* 2018.

139. Battiston, P., Kashyap, R. & Rotondi, V. Trust in Science and Experts during the COVID-19 Outbreak in Italy (2020).

140. Agley, J. Assessing Changes in US Public Trust in Science amid the COVID-19 Pandemic. *Public Health* **183,** 122–125. doi:10.1016/j.puhe.2020.05.004 (2020).

141. Kreps, S. E. & Kriner, D. L. Model Uncertainty, Political Contestation, and Public Trust in Science: Evidence from the COVID-19 Pandemic. *Science Advances* **6,** eabd4563. doi:10.1126/sciadv.abd4563 (2020).

142. Sibley, C. G. *et al.* Effects of the COVID-19 Pandemic and Nationwide Lockdown on Trust, Attitudes toward Government, and Well-Being. *American Psychologist* **75,** 618. doi:10.1037/amp0000662 (2020).

143. Wissenschaft im Dialog. *Science Barometer Special Edition on Corona* https://www.wissenschaft-im-dialog.de/en/our-projects/science-barometer/science-barometer-special-edition-on-corona/. 2020.

144. Open Knowledge Foundation. *Brits Demand Openness from Government in Tackling Coronavirus* https://blog.okfn.org/2020/05/05/brits-demand-openness-from-government-in-tackling-coronavirus/. 2020.

145. Funk, C., Kennedy, B. & Johnson, C. *Trust in Medical Scientists Has Grown in U.S., but Mainly Among Democrats* tech. rep. (2020).

146. Gervais, W. M. *et al.* Global Evidence of Extreme Intuitive Moral Prejudice against Atheists. *Nature Human Behaviour* **1,** 0151. doi:10.1038/s41562-017-0151 (2017).

Acknowledgements

This work was supported by funds from the Templeton Foundation [grant number 60663] to MvE, the Cogito Foundation [grant number R10917] to RMcK, the Australian Research Council [grant number DP180102384] to NL and RMR, Templeton Religion Trust [reference TRT0196] to JAB, and the French Agence Nationale de la Recherche [reference 17-EURE-0017 FrontCog and 10-IDEX-0001-02 PSL] to SA. The analysis was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

Author Contributions

MvE and SH conceptualized the idea, designed the study, and formulated the hypotheses. SA, TB, RB, AC, CG, RG, KH, CK, RMcK, AN, LQ, AR, JER, RMR, HT, FU, RW, DX, and SH provided cultural knowledge (including translations) for adjusting the materials to the national context and collected the data. SH analyzed the data with input from JAB and JMH. SH wrote the first draft of the manuscript, with major critical input from JMH, JAB, RMR, RMcK, and MvE and additional suggestions from all other authors.

Appendix

Hypothesis 3: Cultural Norms Effect

In our preregistration, we formulated a hypothesis about the interaction between source and perceived cultural norms of religiosity in one's country. We expected that this interaction-effect at a country-level would mirror the individual religiosity effect; the relative difference in credibility for the guru's versus the scientist's statement was expected to vary with perceived cultural norms of religiosity per country, i.e., the extent to which religiosity is considered normative and desirable in a society. However, when writing the manuscript we realized that there is no theoretical justification for why perceived religious norms would influence the relative credibility judgment for the two sources, beyond any individual religiosity effect. Furthermore, the way the cultural norms predictor was operationalized in the preregistration was suboptimal; we intended to create an aggregated rating of perceived religious norms at the country level, resulting in only 24 unique values, eliminating all within-country variability and thus greatly reducing the resolution of the data. Using the individual data points would effectively test the hypothesis that "the extent to which I perceive the average citizen in my country to value religion influences my relative credibility evaluation for the scientist vs. the guru, irrespective of my own religious beliefs." We decided that this was in fact an unlikely hypothesis. Nevertheless, we report the results of these suboptimal hypothesis tests here.

Cultural norms of religiosity were measured with two items assessing participants' perception of the extent to which the average person in their country considers a religious lifestyle and belief in God/Gods/spirits important[132]. The preregistration mentioned that responses for the cultural norms variable would be averaged per country to reflect the average perceived cultural norm of religiosity in each country. However, we decided against averaging because that would compromise the informativeness of the data and eliminate the possibility to draw conclusions about whether participants' perception of the cultural norms of religiosity affects their evaluation of the credibility for the statement of the scientist and guru. Note that using the averaged data makes the evidence weaker but does not qualitatively change the results. The presentation order for the personal and cultural norms of religiosity was counterbalanced between participants, to eliminate the possibility for unidirectional anchoring effects. See Table A5 for the exact items and response options.

For hypothesis 3, the model comparison shows that the data provide most evidence for the *null model* that does not include an interaction between source and perceived cultural norms of religiosity, $BF_{10} = 0.04$; $BF_{01} = 22.78$; $BF_{0u} = 73874$. The posterior probability that the interaction is positive across *all* countries is $<.001$; the posterior probability that the *overall* (i.e., the common) interaction effect is positive is 0.63. The mean of the unstandardized source-by-cultural norms of religiosity interaction effect is -0.01 [-0.09, 0.07] and the standard deviation between countries is 0.06.

Additional Model Statistics

For each of the models included in the analyses, we calculated the intraclass correlation (ICC; proportion of the total variance that is accounted for by the clustering) and the explained

Table A1: Bayes factor model comparisons to test $\mathcal{H}_3$

Model		Bayes factor	$p(\mathcal{M})$
$\mathcal{M}_0$	$\text{Country}_u + \text{Source}_u + \text{Norms}_1$	$*$	.96
$\mathcal{M}_1$	$\text{Country}_u + \text{Source}_u + \text{Norms}_1 + \text{Source*Norms}_1$	1-to-22.78	.04
$\mathcal{M}_+$	$\text{Country}_u + \text{Source}_u + \text{Norms}_1 + \text{Source*Norms}_+$	1-to-10^8	$< .01$
$\mathcal{M}_u$	$\text{Country}_u + \text{Source}_u + \text{Norms}_1 + \text{Source*Norms}_u$	1-to-73874	$< .01$

Note. Asterisks mark the preferred model for each hypothesis. The remaining values are the Bayes factors for the respective model vs. the preferred model. Subscripts reflect parameter constraints; $_u$ indicates an unconstrained effect, $_1$ indicates a common (positive/negative) effect, $_+$ indicates a varying positive/negative effect. $p(\mathcal{M})$ gives the posterior model probability. All models include a varying effect of religiosity, a common effect of the source-by-religiosity interaction, and a common effect of the covariate level of education.

variance (Bayesian R^2; proportion of the total variance that is accounted for by the effects). Explained variance was assessed using the `bayes_R2` function from the `rstantools` package[133], based on the method described by Gelman et al.[134]. Explained variance is given separately for general R^2 (all common and varying effects included in the respective model) and for the marginal R^2 (the common effects only). The means and 95% credible intervals for each of the relevant models described in the main text are given in Table A2.

brms Models

Following our preregistration, we also fitted the models in the `brms` R package[131]. For hypotheses 1 (main effect of source) and 2 (interaction between source and individual religiosity) the models fitted in brms corroborated the results from the BayesFactor analyses.

Research Question 1

We preregistered to compare a multilevel ordered probit model with a varying intercept for country[1] to the model that additionally included a common (i.e., fixed) effect of source. The analysis gave a Bayes factor of 4.83×10^{188}, again indicating that credibility rating were higher for the scientist compared to the guru.

Research Question 2

To test the fit effect that one's worldview affects the difference in credibility ratings for the scientist and the guru, we preregistered to compare two models with vs. without an interaction between source and religiosity. The null model was specified as a multilevel ordered probit model with a varying intercept for country and common effects for source and individual religiosity. The alternative model additionally included a common interaction between source and religiosity. Note that in the preregistration, we mentioned that the interaction term should be positive,

[1] We also included a varying intercept for subject, but with only 2 observations per subject fitting a separate intercept for every participant does not make much sense, vastly increases processing time and induces convergence issues. We therefore omitted the varying intercept for subjects.

Table A2: Explained variance and intraclass correlation for all relevant models.

	R^2		Marginal R^2		Intraclass correlation	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Common Effect Models						
Source Effect	0.173	[0.165, 0.182]	0.076	[0.060, 0.094]	0.125	[0.079, 0.198]
Source-by-Religiosity	0.181	[0.172, 0.190]	0.081	[0.062, 0.102]	0.142	[0.095, 0.213]
Processing Time	0.107	[0.099, 0.114]	0.015	[0.012, 0.020]	0.147	[0.091, 0.235]
Memory Performance	0.098	[0.090, 0.105]	0.004	[0.002, 0.006]	0.128	[0.078, 0.207]
Source Effect Trust	0.229	[0.226, 0.232]	0.141	[0.139, 0.143]	0.110	[0.089, 0.134]
Source-by-Religiosity Trust	0.281	[0.278, 0.284]	0.133	[0.110, 0.157]	0.293	[0.258, 0.332]
Varying Effects Models						
Source Effect	0.179	[0.170, 0.187]	0.077	[0.058, 0.099]	0.150	[0.103, 0.220]
Source-by-Religiosity	0.182	[0.174, 0.191]	0.082	[0.064, 0.101]	0.141	[0.095, 0.212]
Processing Time	0.108	[0.100, 0.115]	0.015	[0.011, 0.020]	0.152	[0.097, 0.238]
Memory Performance	0.099	[0.091, 0.106]	0.004	[0.002, 0.006]	0.134	[0.085, 0.210]
Source Effect Trust	0.281	[0.278, 0.283]	0.133	[0.110, 0.157]	0.296	[0.261, 0.334]

Note. Explained variance, split into general explained variance and marginal explained variance (fixed effects only), and intraclass correlations. The 95% CI gives the lower and upper bound of the credible interval. Note that there was no varying effect of the source-by-religiosity interaction for the trust model (validation dataset).

rather than negative. As it concerns an interaction between a continuous variable that has a natural order (low religiosity → high religiosity) and one that has an arbitrary order (guru → scientist or scientist → guru), the sign of interaction term depends entirely on the choice for the reference category. As we believe it is more intuitive to talk about an increase in credibility for the scientist vs. the guru, we used the guru as the reference category. Importantly, the change in sign for the interaction term does thus not reflect a deviation from the preregistered hypotheses. The Bayes factor for the comparison indicated strong evidence in favour of the interaction model: $\mathrm{BF}_{10} = 5.42 \times 10^{22}$.

Research Question 3

In order to test if the worldview-fit effect is also reflected at the country-level, we replaced the individual religiosity predictor in the models for H2 by cultural norms of religiosity. Again, two models were compared with the inclusion of a source*cultural norms interaction as the critical difference between models. As opposed to the results from the BayesFactor models, the brms analysis provides evidence in favor of the source*norms interaction: $\mathrm{BF}_{10} = 67.01$. Importantly, when we added background variables (gender, age, and education) and varying effects of source per country as in the BayesFactor models in Table A1, the evidence for the source*norms interaction disappeared: $\mathrm{BF}_{10} = 0.401$. This suggests that based on the current data, if there is an effect of cultural norms of religiosity on the source credibility effect for a scientist vs a guru, it is at least fragile and small ($\beta = -0.06$, 95% CI$[-0.09, -0.03]$). The individual religiosity effect, on the other hand, appears much more robust.

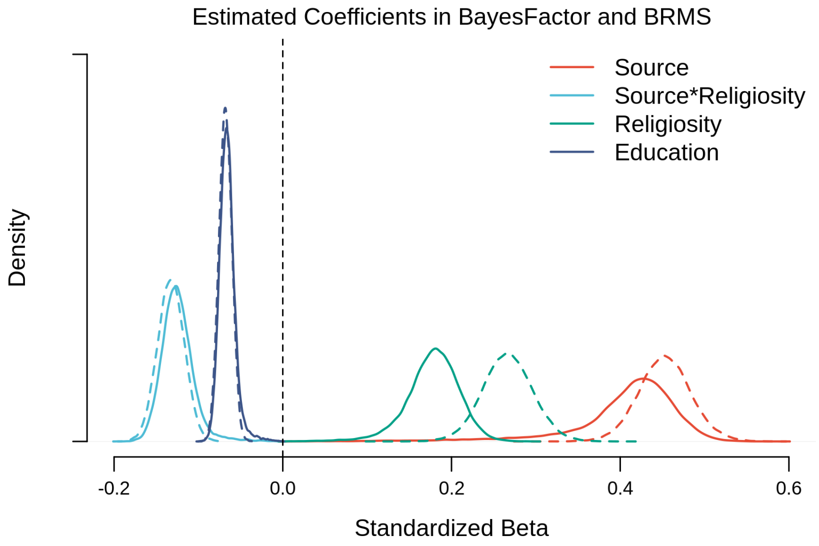


Figure A1: Multilevel estimates of the standardized effects for all included predictors in the unconstrained model for $\mathcal{H}_2$. The solid lines denote density distributions estimated with the `BayesFactor` package[76] and the dashed lines denote the estimations from the `brms` package[131]. The comparison shows that the estimates largely coincide, although the BayesFactor estimates are slightly more conservative, especially for the source effect and the religiosity effect. Note that these two predictors were included as varying effects in the models for both packages.

Table A3: Full model estimates

	BayesFactor Model		brms Model	
	Est.	95% CI	Est.	95% CI
Source	0.407	[0.224, 0.493]	0.453	[0.392, 0.516]
Source*Religiosity	-0.125	[-0.157, -0.081]	-0.133	[-0.163, -0.102]
Religiosity	0.178	[0.108, 0.234]	0.267	[0.208, 0.325]
Education	-0.066	[-0.082, -0.044]	-0.068	[-0.083, -0.053]

Note. Est. = estimate; CI = credible interval. Estimates are standardized parameter estimates from the full model for $\mathcal{H}_2$ as reported in the main text and its ordinal equivalent in brms[131].

Comparison estimates in BayesFactor and brms

Finally, in addition to the derived Bayes factors, we also compared the estimates of the best-fitting model from the BayesFactor model to those from the brms model. This concerns the model with varying effects for gender, age, education, source, and religiosity and a common effect for the source*religiosity interaction. In brms the parameters are automatically standardized for ordinal regression using the cumulative probit link function. Therefore, we also standardized the parameters in the BayesFactor models (by standardizing the data, including the outcome variable). As shown in Figure A1 and Table A3 the estimates for the included predictors are largely similar, with slightly more conservative estimates for the BayesFactor model. The main effect of religiosity seems the only estimate that is substantially smaller in the normal BayesFactor models compared to the ordinal brms models.

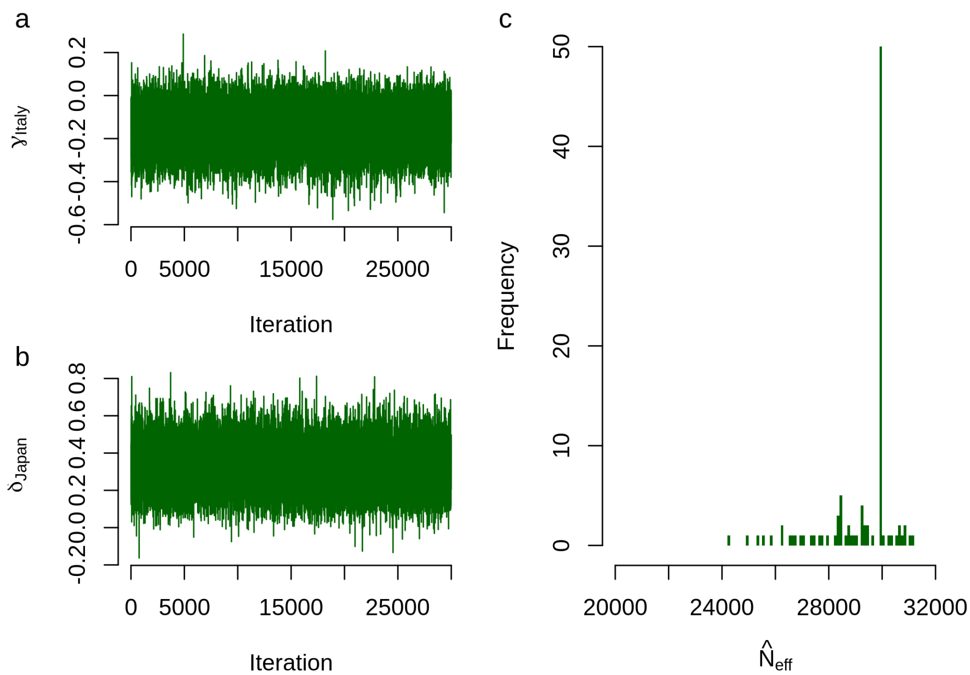


Figure A2: MCMC diagnostics. **a.** Chains for parameters with the smallest (varying slope for source effect in Italy) and **b.** largest (varying slope for the religiosity effect in Japan) $\hat{R}$ values. **c.** Number of effective samples for each parameter in the full model.

MCMC Diagnostics

To investigate convergence of the MCMC chains, we calculated split-$\hat{R}$[135] based on the rank-based method described in Vehtari et al.[136]. The smallest and largest $\hat{R}$ values were 0.99997 and 1.00040, respectively, indicating good within-chain convergence. The traceplots for these smallest and largest $\hat{R}$ values are shown in Figure A2a and b.

The number of effective samples ($\hat{N}_{eff}$) was calculated per parameter to assess to what extent autocorrelation in the chains reduces the certainty of the posterior estimates[137]. Ideally, $\hat{N}_{eff}$ is as large as possible[136]. The $\hat{N}_{eff}$ for each of the 107 estimated parameters is displayed in Figure A2c. Note that $\hat{N}_{eff}$ can be larger than the the total number of iterations (in this case: $N = 30,000$) when the samples are anti-correlated or antithetical[138]. The smallest $\hat{N}_{eff} = 24,210.67$ for the varying slope of the source-by-religiosity interaction for Croatia. For many parameters, $\hat{N}_{eff}$ is equal to the number of iterations or even higher. We therefore concluded that the effective sample size is sufficient for valid interpretation of the estimates and inference.

Country Comparisons Across Datasets

To explore the country-level patterns in the source effect between both datasets, we assessed the correlation between the experimental source credibility effect in the primary dataset and the contrast of the trust ratings for scientists and traditional healers in the validation dataset per country. The raw observed relation as well as the relation between the modeled source effects are depicted in Figure A3a and b. The plots do not suggest a strong correlation between source effects, which is corroborated by the evidence for the correlation: $\text{BF}_{+0} = 1.06$; $\text{BF}_{+0} = 0.97$ for the observed and estimated source effects, respectively. These Bayes factors imply "absence of

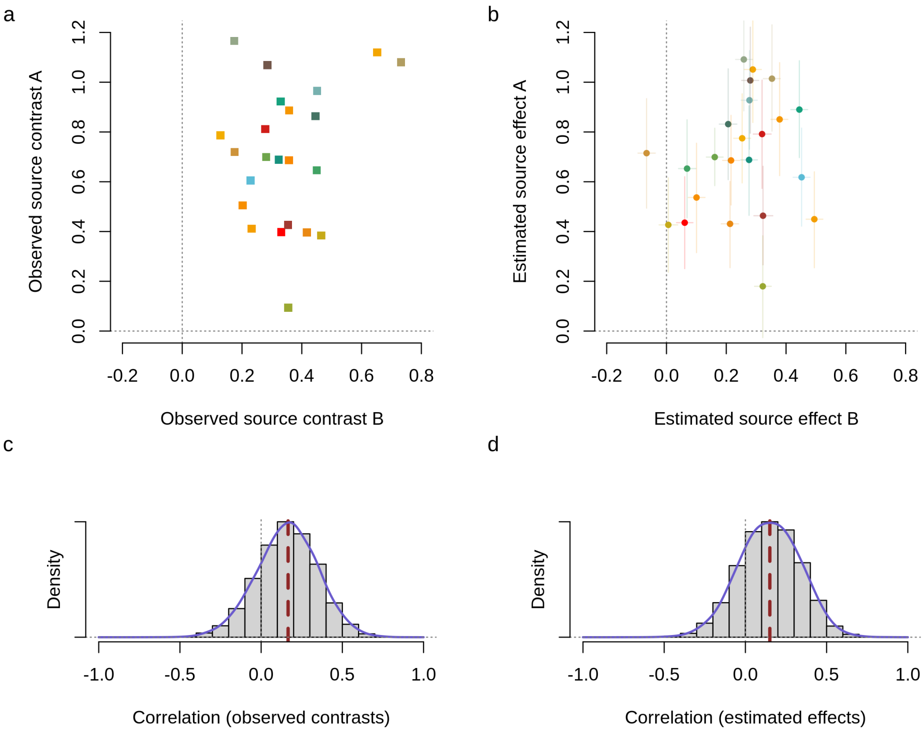


Figure A3: Correlation between the source effect in the new experimental dataset (set A) and the validation survey data on trust (set B). Panel **a** shows the relationship between the observed contrast effects (scientist minus guru) in both datasets. Each square represents a country. Panel **b** shows the country-level estimates (medians) of the source effect in the experimental dataset and the validation dataset. Each dot represents a country. The horizontal and vertical lines denote the 95% credible intervals. Panels **c** and **d** display the posterior distribution of the correlation coefficient ρ using the observed contrasts and estimated effects, respectively. The vertical dashed line reflects the median value for ρ.

evidence", meaning that we cannot conclude whether or not the country-level source effects are related between the two datasets. The 95% credible intervals further support the uncertainty of the correlation: $\rho_{obs} = 0.17$ [-0.22,0.52]; $\rho_{est} = 0.15$ [-0.22,0.50]. We note however, that in addition to the uncertainty related to the small number of observations[2], caution is also warranted due to the difference in included samples and exact items (credibility of specific nonsense statements vs. explicit trust in authorities) between datasets.

Causal Assumptions and Covariate Selection

In order to systematically investigate which third variables should and should not be included in the statistical model, we used graphical causal models representing the relations between the variables in our data. As part of the data of interest is observational (e.g., religiosity, demographics), it is important to identify potential confounder variables, 'back-door paths', mediators and colliders that may affect causal inference[119–121]. We identified the following structure based on theoretical assumptions about the measured variables:

[2]These were the 24 countries from the main dataset minus China, for which no religiosity data was available in the validation dataset.

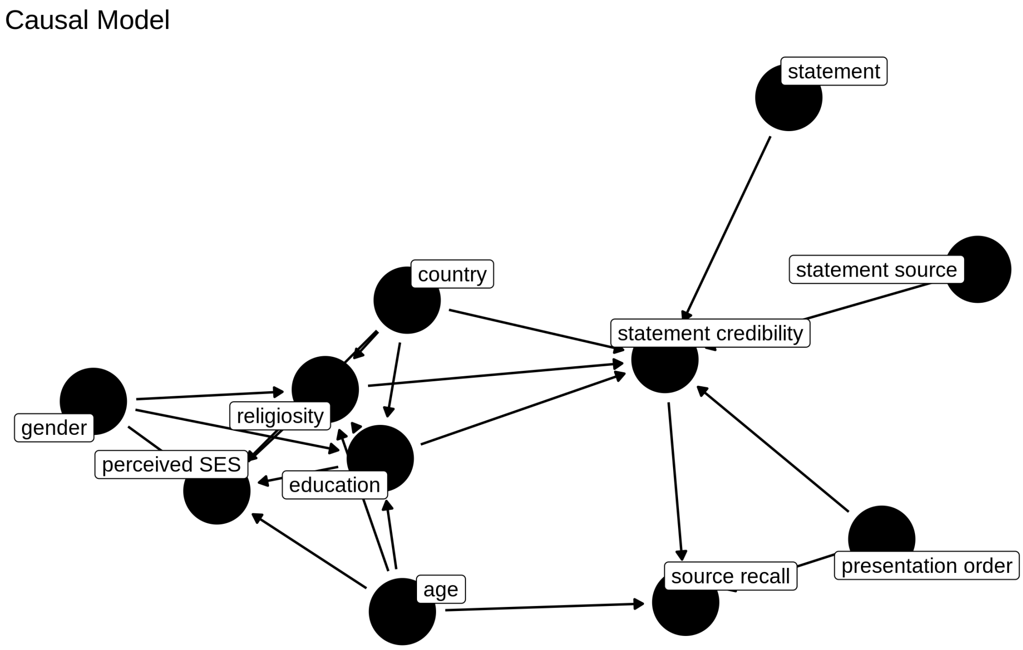


Figure A4: Graphical model for the causal structure of the variables in the data.

- differences in perceived credibility of goobledegook statements are (potentially) affected by:
  - the source of the statement (scientist vs. guru)
  - order of presentation
  - the statement itself
  - country (culture)
  - education (skepticism)
  - religion.
- religion is affected by age, SES, education, gender, and country
- SES is affected by country, education, age, and gender
- education is affected by country, age, and gender
- recall of the source is a function of credibility, age and presentation order

Using *directed acyclic graphs* (DAGs[118]) created in the R package `ggdag`[122], this resulted in the structure as displayed in Figure A4. The adjustment set in Figure A5 shows that assuming this model, we should only condition on (i.e., include) *country* and *education* as covariates or adjustment variables. So, rather than "controlling for" all indicators that could affect either the predictor or outcome of interest, we only adjusted for the indicators that are needed for causal inference. Also note that experimental indicators such as presentation order and statement version were fully counterbalanced between participants. As drawn in Figure A6, in the large model, many covariates are identified as colliders; including those may introduce spurious associations and bias the relation of interest between religiosity and (source) credibility. In the adjusted model, none of the remaining covariates are colliders, making conditioning on country and education valid inference choices.

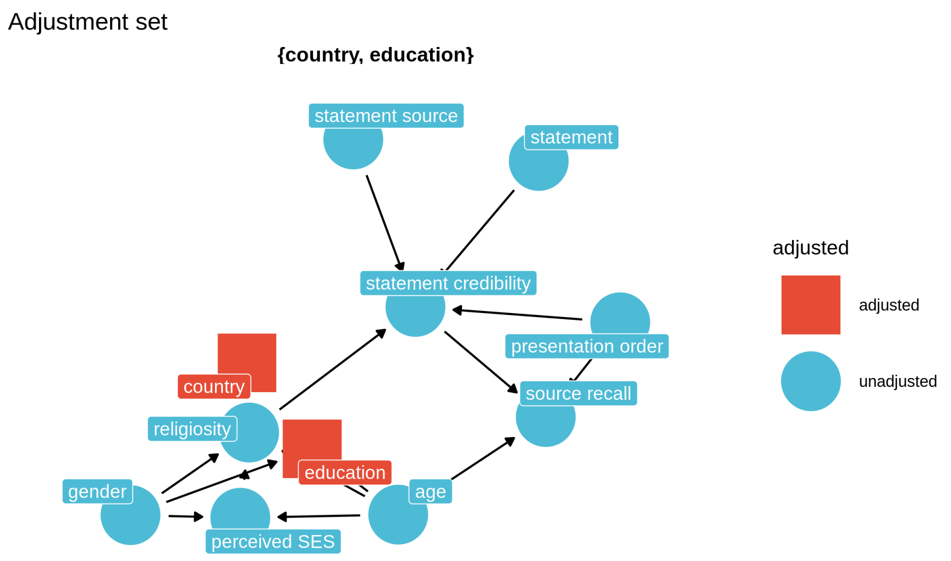


Figure A5: Graphical model of the adjusted set showing which variables (in red) should be conditioned on for valid causal inference.

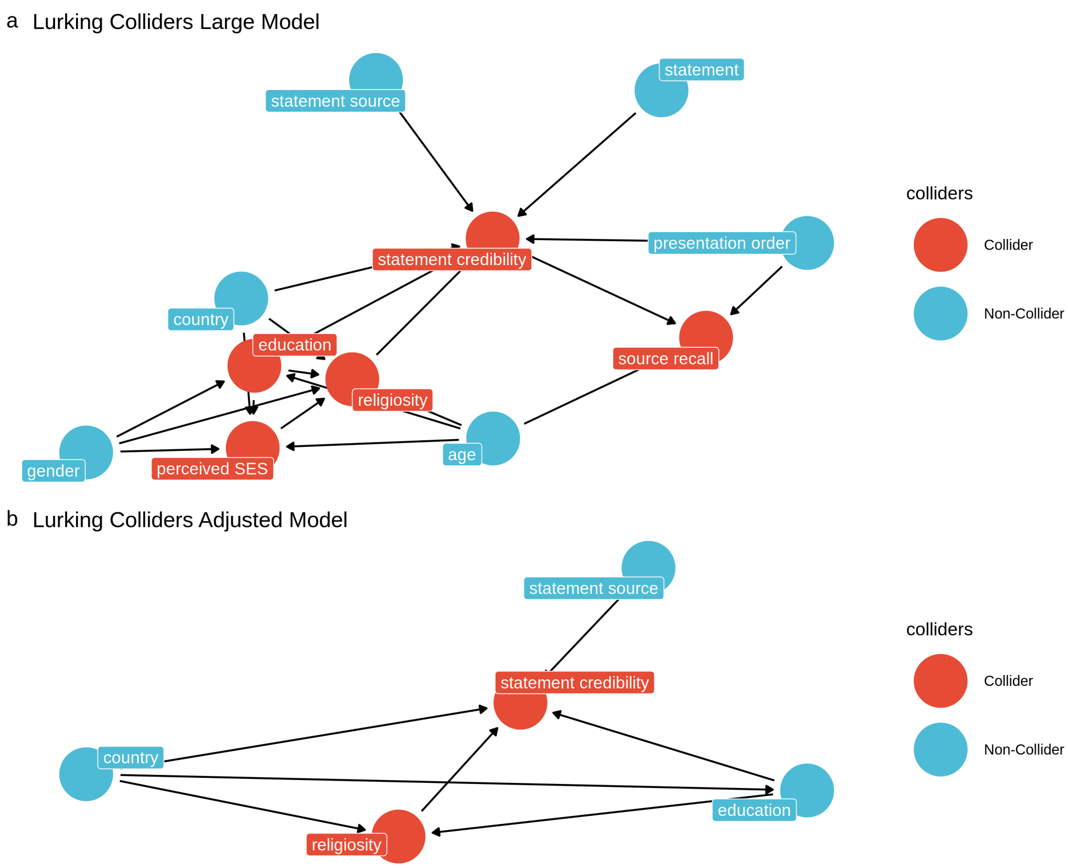


Figure A6: Potential colliders in the causal structure for the (a) large model and the (b) adjusted model.

A Note on Scientific Credibility and the COVID-19 Situation

In the main paper, we included the case of COVID-19 only as a timely example to introduce our general topic, but we do not further elaborate on trust and credibility of authorities related to COVID-19 specifically. That is, we believe that our findings bear a broader and more general relevance for understanding source credibility-effects, that go beyond the current situation. Many others have investigated the perception of experts in relation to COVID-19 specifically in great detail, see for instance[139–145]. While we do not discuss COVID-19 at length in the main paper, we quickly reflect here on the potential implications of these findings, using the Netherlands as an illustration.

The pattern found in the studies referred to above is somewhat mixed, yet most data seem to suggest that trust in science/scientists has either remained the same or even increased during the pandemic. In the Netherlands for instance, the majority of the general public also still places more trust in the Outbreak Management Team (OMT; a team of experts convened to advise the government on policy in the event of an outbreak of infectious disease) and RIVM (Dutch equivalent of the CDC) than Maurice de Hond or Willem Engel (Dutch public figures and self-declared COVID-19 experts). This is for instance indirectly indicated by increased vaccination willingness over the last months (about 80% in NL). Moreover, the public still mostly relies on information regarding vaccination provided by vaccination centers (60.6%), the RIVM website (48.1%) and GPs (39.6%), to a stronger extent than that provided by the media (34.8%), trusted celebrities (2.5%) or social media (2%; see www.rivm.nl/gedragsonderzoek/maatregelen-welbevinden/vaccinatiebereidheid). So while there are certainly individual differences in the perception of who is considered an expert, it seems that, on average, scientific expertise is still considered the most trustworthy source of information compared to other sources in relation to COVID-19 - and perhaps more generally as our study suggests.

Supplementary Tables and Figures

Table A4: Estimates per country

	Intercept		Source Effect		Source*Religiosity	
Country	Est.	95% CI	Est.	95% CI	Est.	95% CI
Total	3.972	[3.747, 4.198]	0.696	[0.598, 0.794]	-0.214	[-0.294, -0.136]
Australia	4.328	[4.222, 4.433]	0.553	[0.366, 0.738]	-0.266	[-0.415, -0.119]
Belgium	3.655	[3.525, 3.786]	0.690	[0.475, 0.908]	-0.286	[-0.496, -0.085]
Brazil	4.191	[4.077, 4.303]	0.558	[0.361, 0.752]	-0.225	[-0.392, -0.058]
Canada	3.941	[3.821, 4.059]	0.930	[0.726, 1.141]	-0.183	[-0.379, 0.011]
Chile	4.116	[3.994, 4.238]	0.785	[0.575, 0.994]	-0.328	[-0.530, -0.131]
China	5.049	[4.940, 5.159]	0.444	[0.246, 0.639]	-0.169	[-0.372, 0.036]
Croatia	3.444	[3.323, 3.564]	0.692	[0.483, 0.898]	-0.006	[-0.185, 0.179]
Denmark	3.494	[3.383, 3.606]	0.821	[0.629, 1.014]	-0.179	[-0.362, 0.002]
France	3.815	[3.705, 3.925]	0.630	[0.434, 0.819]	-0.131	[-0.318, 0.064]
Germany	4.258	[4.198, 4.319]	0.688	[0.573, 0.804]	-0.067	[-0.193, 0.064]
India	4.907	[4.680, 5.134]	0.491	[0.211, 0.760]	-0.299	[-0.515, -0.087]
Ireland	4.010	[3.904, 4.116]	0.535	[0.346, 0.722]	-0.341	[-0.516, -0.168]
Israel	4.095	[4.000, 4.189]	0.766	[0.597, 0.937]	-0.206	[-0.382, -0.034]
Italy	4.078	[3.953, 4.203]	0.967	[0.757, 1.183]	-0.161	[-0.364, 0.044]
Japan	3.912	[3.799, 4.023]	0.424	[0.229, 0.617]	-0.208	[-0.432, 0.016]
Lithuania	3.548	[3.425, 3.671]	0.815	[0.604, 1.029]	-0.244	[-0.453, -0.036]
Morocco	4.053	[3.902, 4.207]	0.628	[0.389, 0.863]	-0.098	[-0.257, 0.065]
Netherlands	3.280	[3.179, 3.382]	0.654	[0.472, 0.831]	-0.127	[-0.296, 0.045]
Romania	4.354	[4.248, 4.460]	0.575	[0.391, 0.758]	-0.276	[-0.444, -0.110]
Singapore	3.904	[3.778, 4.032]	0.754	[0.544, 0.965]	-0.229	[-0.446, -0.014]
Spain	3.474	[3.341, 3.609]	0.895	[0.677, 1.122]	-0.219	[-0.423, -0.015]
Turkey	3.583	[3.470, 3.693]	1.026	[0.825, 1.233]	-0.198	[-0.363, -0.034]
UK	3.682	[3.562, 3.803]	0.769	[0.566, 0.972]	-0.365	[-0.569, -0.169]
US	4.110	[4.001, 4.219]	0.692	[0.503, 0.882]	-0.369	[-0.548, -0.198]

Note. Est. = estimate; CI = credible interval. Estimates are unstandardized parameter estimates from the full model for $\mathcal{H}_2$ as reported in the main text.

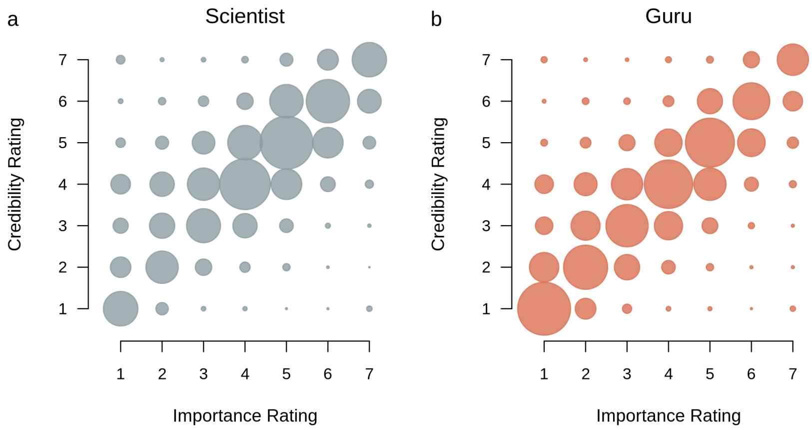


Figure A7: Correlation between the credibility rating and importance rating per source. The size of the bubbles reflects the relative number of observations for the respective value on the discrete scale.

Religiosity Items

Table A5: Religiosity Items

Individual Religiosity	
1.	Apart from weddings and funerals, about how often do you attend religious services these days? [Never, practically never – more than once a week] (7-pt)
2.	How often do you pray/meditate? [Never, practically never – several times a day] (8-pt)
3.	Independently of whether you attend religious services or not, would you say you are: [A religious person / not a religious person / an atheist]
4.	Do you belong to a religion or religious denomination? If so, which one? [Yes / No, *options tailored to respective country*]
5.	To what extent do you believe in God? [Not at all – very much] (7-pt)
6.	To what extent do you believe in life after death? [Not at all – very much] (7-pt)
7.	In your life, how important is a religious lifestyle? [Not at all important – extremely important] (5-pt)
8.	In your life, how important is belief in God? [Not at all important – extremely important] (5-pt)
Cultural Norms of Religiosity	
9.	For an average US*citizen, how important would you say is a religious lifestyle? [Not at all important – extremely important] (5-pt)
10.	For an average US*citizen, how important would you say is belief in God? [Not at all important – extremely important] (5-pt)

Note. Labels for the response options are given in square brackets, with the number of Likert scale options in round brackets (where applicable). The differences in range of the response scales are inherent to the fact that they are taken from existing scales. As we wanted to stay as close to the original scales as possible, we refrained from modifying the response options.
* Adjusted to the nationality of each country.

Religious Replication Project

The aim of the religious replication project is to establish the robustness and potential boundary conditions of classical findings in the psychology and cognitive science of religion. To this end we conducted a large cross-cultural study by using standardized surveys and tasks in different countries (for a similar approach, see[78,146]). We focused on four related topics: (1) the relation between religion and well-being, (2) the effects of religious and non-religious displays on perceived trustworthiness, (3) effects of source credibility on the perception of pseudo-profound statements, and (4) dualist thinking and religion. These topics were combined in one package, consisting of different scales and experimental manipulations. The current study focuses on the the third sub-study, preregistration documents for the other three can also be found on the OSF (osf.io/dj6ck/).