nature portfolio

https://doi.org/10.1038/s41562-021-01273-8

Supplementary information

The Einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity

In the format provided by the authors and unedited

Supplementary Information

"The Einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity"

Suzanne Hoogeveen*, Julia M. Haaf, Joseph A. Bulbulia, Robert M. Ross, Ryan McKay, Sacha Altay, Theiss Bendixen, Renatas Berniūnas, Arik Cheshin, Claudio Gentili, Raluca Georgescu, Will M. Gervais, Kristin Hagel, Christopher Kavanagh, Neil Levy, Alejandra Neely, Lin Qiu, André Rabelo, Jonathan E. Ramsay, Bastiaan T. Rutjens, Hugh Turpin, Filip Uzarevic, Robin Wuyts, Dimitris Xygalatas, & Michiel van Elk

This document contains supplementary information for the manuscript "The Einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity". Specifically, it provides additional methodological details (i.e., religiosity items) and additional results (i.e., results for the third hypothesis, additional details on statistical models, all estimates per country, convergence, and causal assumptions, results using an alternative statistics package, and a country-level comparison across the two different datasets). Furthermore, we included a note on the relevance for the current work and the covid19 situation and a brief description of the overall Religious Replication Project.

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[1]. 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

*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 and analysis code are provided at https://osf.io/qsyvw/.

Supplementary Table 1: Bayes factor model comparisons to test $\mathcal{H}_3$

Model		Bayes factor	$p(\mathcal{M})$
$\mathcal{M}_0$	Country_u + Source_u + Norms_1	*	.96
$\mathcal{M}_1$	Country_u + Source_u + Norms_1 + Source*Norms_1	1-to-22.78	.04
$\mathcal{M}_+$	Country_u + Source_u + Norms_1 + Source*Norms_+	1-to-10^8	< .01
$\mathcal{M}_u$	Country_u + Source_u + Norms_1 + 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.

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 Supplementary Table 5 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, $\text{BF}_{10} = 0.04$; $\text{BF}_{01} = 22.78$; $\text{BF}_{0u} = 73874$. 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 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[2], based on the method described by Gelman et al.[3]. 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 Supplementary Table 2.

brms Models

Following our preregistration, we also fitted the models in the `brms` R package[4]. 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.

Supplementary Table 2: 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).

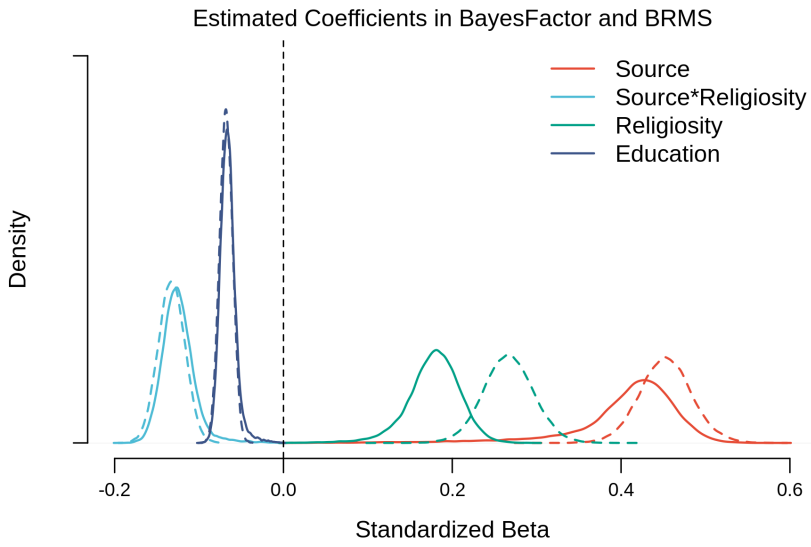
Research Question 1

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, rather than negative. As it concerns an interaction between a continuous variable that has a natural order (low religiosity $\rightarrow$ high religiosity) and one that has an arbitrary order (guru $\rightarrow$ scientist or scientist $\rightarrow$ 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: $\text{BF}_{10} = 5.42 \times 10^{22}$.

[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.



Supplementary Figure 1: 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[5] and the dashed lines denote the estimations from the `brms` package[4]. 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.

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: $\text{BF}_{10} = 67.01$. Importantly, when we added background variables (gender, age, and education) and varying effects of source per country as in the BayesFactor models in Supplementary Table 1, the evidence for the source*norms interaction disappeared: $\text{BF}_{10} = 0.401$. 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 ($\beta = -0.06,\ 95\%\ \text{CI}[-0.09, -0.03]$). The individual religiosity effect, on the other hand, appears much more robust.

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 Supplementary Figure 1 and Supplementary Table 3 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.

Supplementary Table 3: 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[4].

MCMC Diagnostics

To investigate convergence of the MCMC chains, we calculated split-$\hat{R}$[6] based on the rank-based method described in Vehtari et al.[7]. 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 Supplementary Figure 2a 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[8]. Ideally, $\hat{N}_{eff}$ is as large as possible[7]. The $\hat{N}_{eff}$ for each of the 107 estimated parameters is displayed in Supplementary Figure 2c. 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[9]. 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.

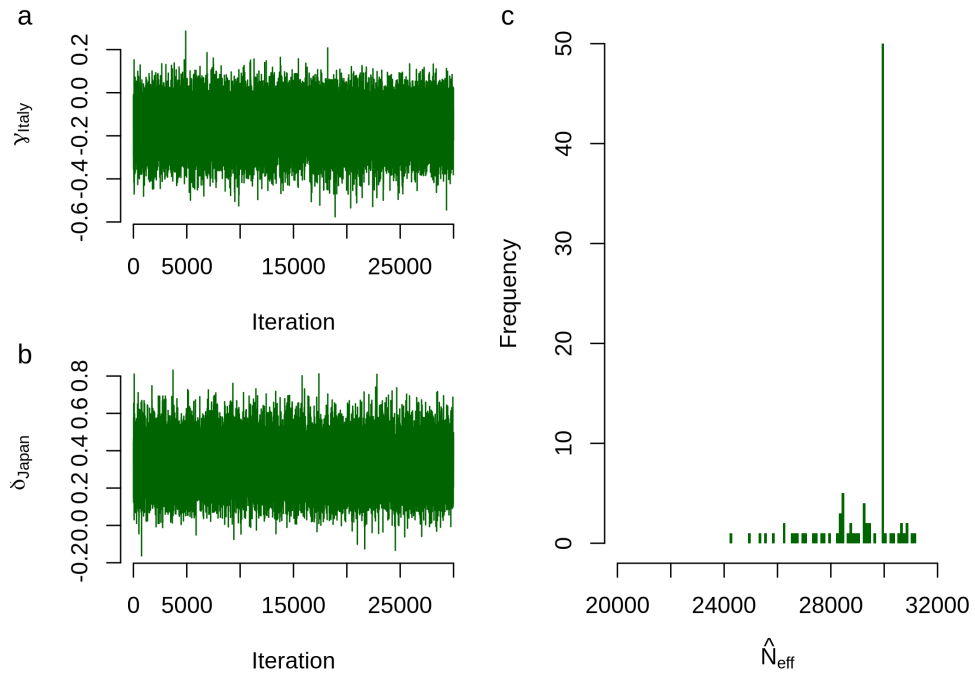
Country Comparisons Across Datasets

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 Supplementary Figure 3a and b. The plots do not suggest a strong correlation between source effects, which is corroborated by the evidence for the correlation: $BF_{+0} = 1.06$; $BF_{+0} = 0.97$ for the observed and estimated source effects, respectively. These Bayes factors imply "absence of 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

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

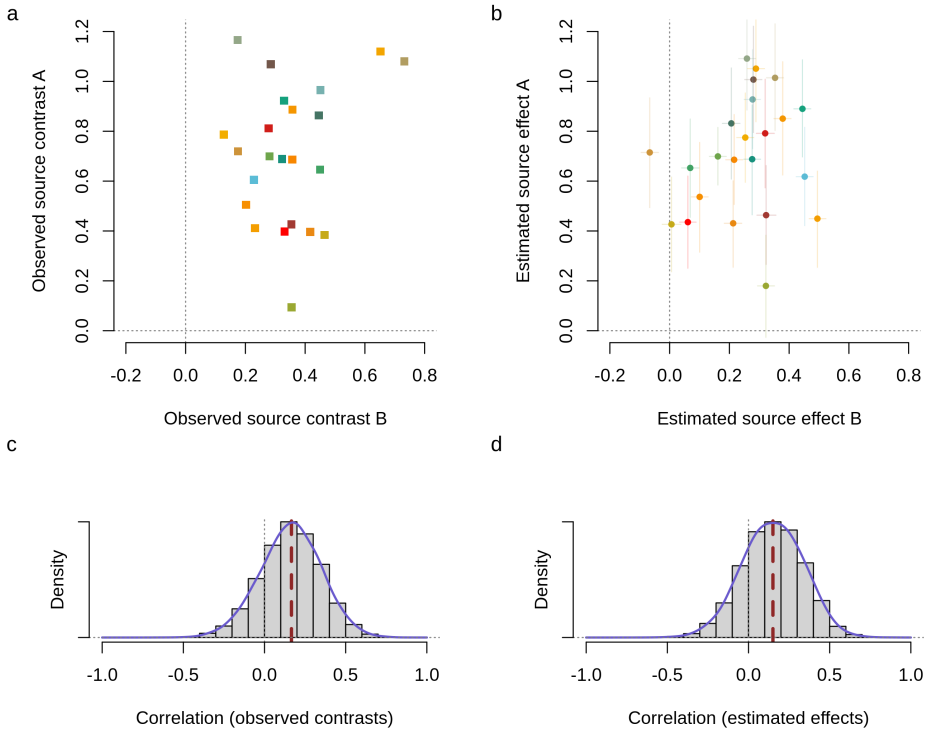


Supplementary Figure 2: MCMC diagnostics. **a.** Chains for parameters with the smallest (varying slope for source effect in Italy) and **b.** largest (varying slope for the religiosity effect in Japan) $\hat{R}$ values. **c.** Number of effective samples for each parameter in the full model.

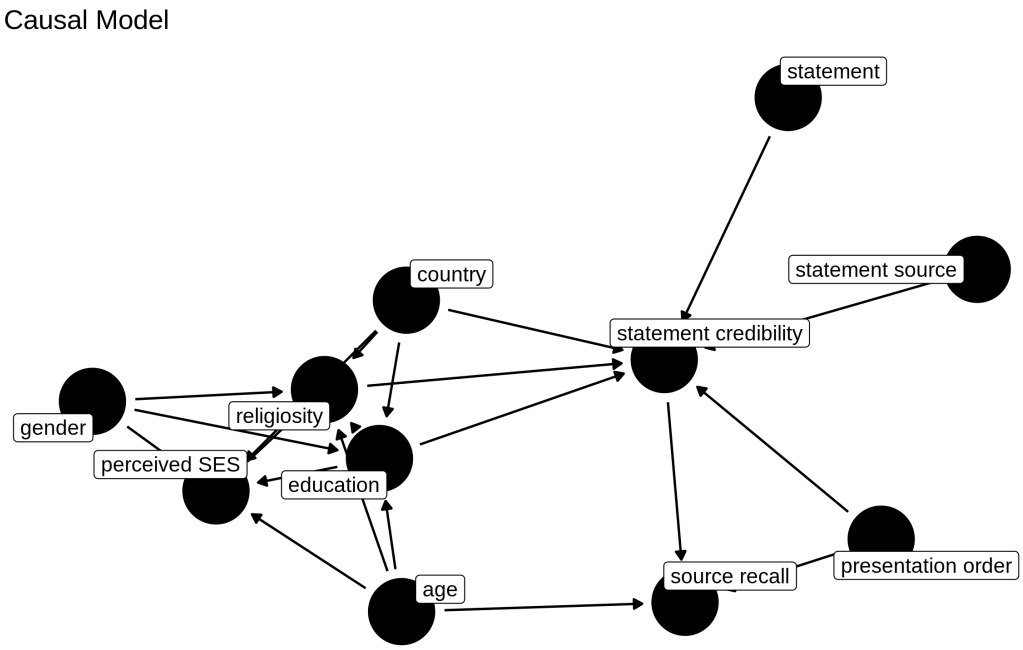
variables in our data. As part of the data of interest is observational (e.g., religiosity, demographics), it is important to identify potential confounder variables, 'back-door paths', mediators and colliders that may affect causal inference[10–12]. We identified the following structure based on theoretical assumptions about the measured variables:

- 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[13]) created in the R package `ggdag`[14], this resulted in the structure as displayed in Supplementary Figure 4. The adjustment set in Supplementary Figure 5 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 Supplementary Figure 6, in the large model, many covariates are identified as colliders; including those may introduce spurious associations and bias the relation of interest between religiosity



Supplementary Figure 3: 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.



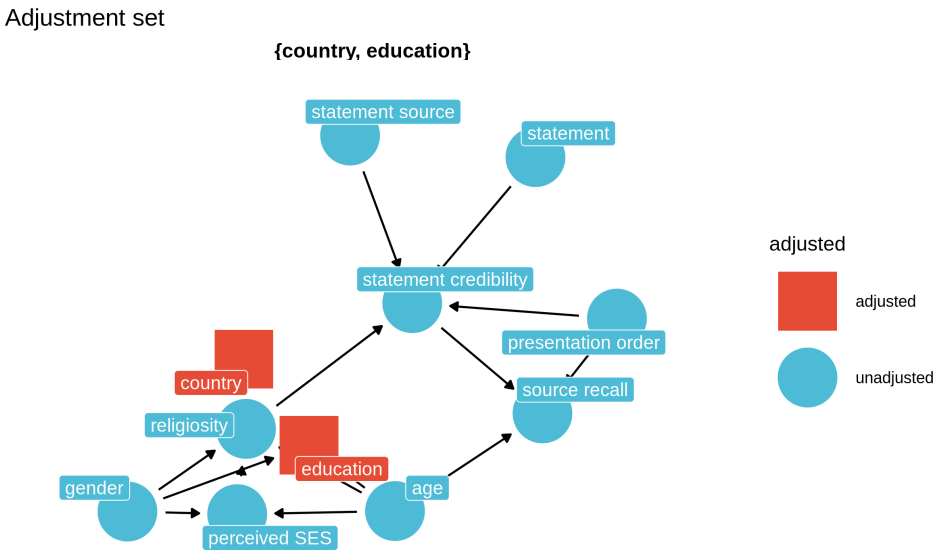
Supplementary Figure 4: Graphical model for the causal structure of the variables in the data.

and (source) credibility. In the adjusted model, none of the remaining covariates are colliders, making conditioning on country and education valid inference choices.

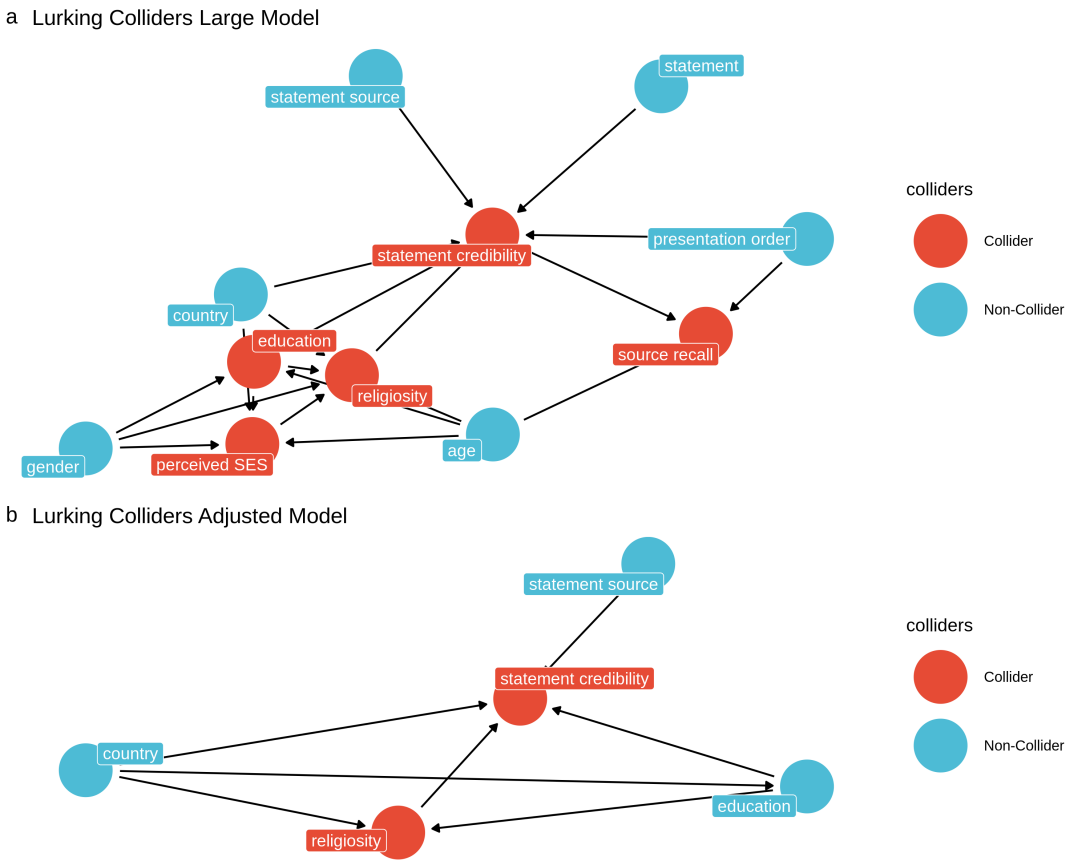
A Note on Scientific Credibility and the COVID-19 Situation

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[15–21]. 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 Figure 5: Graphical model of the adjusted set showing which variables (in red) should be conditioned on for valid causal inference.



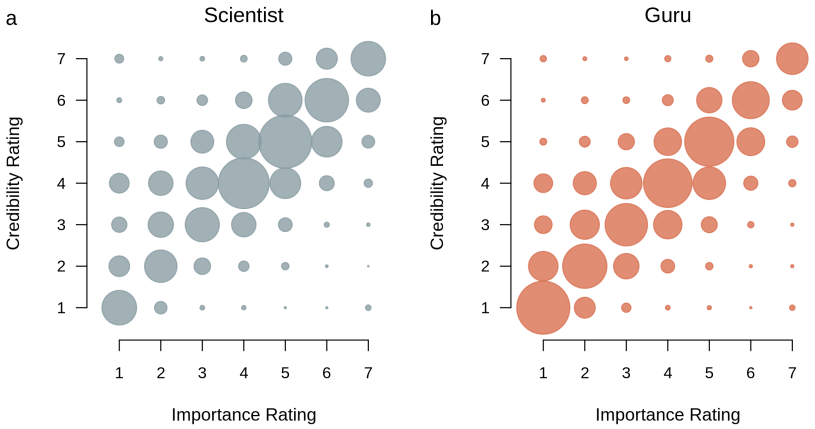
Supplementary Figure 6: Potential colliders in the causal structure for the (a) large model and the (b) adjusted model.

Supplementary Tables and Figures

Supplementary Table 4: 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.



Supplementary Figure 7: 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

Supplementary Table 5: 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[22,23]). 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/).

References

1. 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).
2. Gabry, J., Goodrich, B. & Lysy, M. *Rstantools: Tools for Developing r Packages Interfacing with Stan.* 2020.
3. 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).
4. 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).
5. Morey, R. D. & Rouder, J. N. *BayesFactor: Computation of Bayes Factors for Common Designs* 2018.
6. Gelman, A. *et al. Bayesian Data Analysis (3rd Ed.)* (Chapman & Hall/CRC, Boca Raton (FL), 2014).
7. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC. *Bayesian Analysis,* 1–28. doi:10.1214/20-BA1221 (2021).
8. 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.
9. Carpenter, B. *We Were Measuring the Speed of Stan Incorrectly—It's Faster than We Thought in Some Cases Due to Antithetical Sampling* 2018.
10. 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).
11. 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).
12. 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).
13. Pearl, J. Causal Diagrams for Empirical Research. *Biometrika* **82,** 669–688. doi:10.1093/biomet/82.4.669 (1995).
14. Barrett, M. *Ggdag: Analyze and Create Elegant Directed Acyclic Graphs* 2021.
15. Battiston, P., Kashyap, R. & Rotondi, V. Trust in Science and Experts during the COVID-19 Outbreak in Italy (2020).
16. 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).
17. 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).

18. 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).

19. 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.

20. 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.

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

22. 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).

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