



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

Suzanne Hoogeveen<sup>1</sup>✉, Julia M. Haaf<sup>1</sup>, Joseph A. Bulbulia<sup>2</sup>, Robert M. Ross<sup>3</sup>, Ryan McKay<sup>4</sup>, Sacha Altay<sup>5</sup>, Theiss Bendixen<sup>6</sup>, Renatas Berniūnas<sup>7</sup>, Arik Cheshin<sup>8</sup>, Claudio Gentili<sup>9</sup>, Raluca Georgescu<sup>10</sup>, Will M. Gervais<sup>11</sup>, Kristin Hagel<sup>12</sup>, Christopher Kavanagh<sup>13,14</sup>, Neil Levy<sup>15</sup>, Alejandra Neely<sup>16</sup>, Lin Qiu<sup>17</sup>, André Rabelo<sup>18</sup>, Jonathan E. Ramsay<sup>19</sup>, Bastiaan T. Rutjens<sup>1</sup>, Hugh Turpin<sup>13</sup>, Filip Uzarevic<sup>20</sup>, Robin Wuyts<sup>1</sup>, Dimitris Xygalatas<sup>21</sup> and Michiel van Elk<sup>22</sup>

**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 and all levels of religiosity, scientists held greater authority than spiritual gurus. In addition, individual religiosity predicted a weaker relative preference for the statement from the scientist compared with the spiritual guru, and was more strongly associated with credibility judgements for the guru than the scientist. Independent data on explicit trust ratings across 143 countries mirrored our experimental findings. These findings suggest that irrespective of one's religious worldview, across cultures science is a powerful and universal heuristic that signals the reliability of information.**

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'<sup>1</sup>. A group of infectious disease experts and former colleagues from Stanford, however, publicly criticized Dr Atlas, who is a radiologist, for spreading 'falshoods and misrepresentation of science' through his statements about face masks, social distancing and the safety of community transmission<sup>2</sup>. In the 2020 pandemic crisis, all eyes turned to scientific experts to provide advice, guidelines and remedies; from COVID-19 alarmists to sceptics, appeal to scientific authority appeared a prevalent strategy on both sides of the political spectrum. Please see the Supplementary Information for a short commentary on how the current 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<sup>3–10</sup>. Children and adults are sensitive to the past track record of informants<sup>11–16</sup>, evidence of their benevolence toward the recipient of testimony<sup>17–19</sup>, as well as how credible the information is at

face value<sup>20,21</sup>. 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<sup>22–28</sup>. 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<sup>29</sup>, 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

<sup>1</sup>Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands. <sup>2</sup>School of Psychology, Victoria University of Wellington, Wellington, New Zealand. <sup>3</sup>Department of Psychology, Macquarie University, Sydney, New South Wales, Australia. <sup>4</sup>Department of Psychology, Royal Holloway, University of London, Egham, UK. <sup>5</sup>Département d'études cognitives, Institut Jean Nicod, Paris, France. <sup>6</sup>Department of the Study of Religion, Aarhus University, Aarhus, Denmark. <sup>7</sup>Institute of Psychology, Vilnius University, Vilnius, Lithuania. <sup>8</sup>Department of Human Services, University of Haifa, Haifa, Israel. <sup>9</sup>Department of General Psychology, University of Padova, Padua, Italy. <sup>10</sup>International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babes-Bolyai University, Cluj-Napoca, Romania. <sup>11</sup>Centre for Culture and Evolution, Brunel University London, Uxbridge, UK. <sup>12</sup>Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>13</sup>School of Anthropology and Museum Ethnography, University of Oxford, Oxford, UK. <sup>14</sup>Department of Psychology, Rikkyo University, Tokyo, Japan. <sup>15</sup>Department of Philosophy, Macquarie University, Sydney, New South Wales, Australia. <sup>16</sup>School of Psychology, Adolfo Ibáñez University, Santiago, Chile. <sup>17</sup>Department of Psychology, Nanyang Technological University, Singapore, Singapore. <sup>18</sup>Department of Social and Work Psychology, Universidade de Brasília, Brasília, Brazil. <sup>19</sup>School of Social and Health Sciences, James Cook University, Singapore, Singapore. <sup>20</sup>Department of Psychology, Catholic University of Louvain, Louvain-la-Neuve, Belgium. <sup>21</sup>Department of Anthropology and Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA. <sup>22</sup>Institute of Psychology, Leiden University, Leiden, The Netherlands.

✉e-mail: [suzanne.j.hoogeveen@gmail.com](mailto:suzanne.j.hoogeveen@gmail.com)

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<sup>30,31</sup> and scientific information is often difficult and counterintuitive<sup>32–34</sup>. 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<sup>9,10,35</sup>. 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<sup>36</sup>. Some experimental evidence, for instance, suggests that irrelevant neuroscience information<sup>37–39</sup> or nonsense mathematical equations<sup>40</sup> can boost the perceived quality of presented claims, though note that replication studies suggest that mere brain images may not suffice<sup>41,42</sup>. Notably, these effects were present only among non-experts (that is, 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, previous beliefs about religious or spiritual texts instigate expectations that the information presented will be obscure. Supernatural explanations often appeal to phenomena that operate outside the natural world and to experiences deemed ineffable, mysterious and exempt from empirical validation<sup>43–48</sup>. 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<sup>49,50</sup>. 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 depends on individual difference factors such as the perceiver’s (political) ideology and worldview<sup>51–54</sup>. In the absence of the means to rationally evaluate a claim and reliable source information, people probably infer credibility based on beliefs about the group to which the source belongs (for example, ‘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<sup>23,55</sup>. 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<sup>56</sup> and to require less evidence for religious claims (for example, efficacy of prayer to cure illness) than for scientific claims (for example, efficacy of medication<sup>57,58</sup>). 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<sup>59</sup>. This effect was moderated by the amount of contact participants had with the specific group to which the religious leader belonged, highlighting the importance of the person–source fit for message acceptance.

To account for these effects, alongside traditional dual-process models of persuasion<sup>9,10,60,61</sup>, various authors have recently proposed a Bayesian framework in which subjective beliefs about the source (for example, trustworthiness) and one’s worldviews contribute to belief updating in response to new information following Bayesian principles<sup>6,62–64</sup>. By including background beliefs, these Bayesian networks describe how a differential weighing of evidence and even divergent updating (belief polarization) can be considered rational

**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		
$\mathcal{M}_0$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub>	1-to-10 <sup>228</sup>	<0.01
$\mathcal{M}_1$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub>1</sub>	1-to-10 <sup>17</sup>	<0.01
$\mathcal{M}_+$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub>+</sub>	*	0.92
$\mathcal{M}_u$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub><i>u</i></sub>	1-to-12.30	0.08
Hypothesis 2: Source-by-religiosity effect		
$\mathcal{M}_0$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub><i>u</i></sub>	1-to-10 <sup>15</sup>	<0.01
$\mathcal{M}_1$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub><i>u</i></sub> + Source × Religiosity <sub>1</sub>	*	0.50
$\mathcal{M}_+$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub><i>u</i></sub> + Source × Religiosity <sub>+</sub>	1-to-1.28	0.39
$\mathcal{M}_u$ Country <sub><i>u</i></sub> + Religiosity <sub><i>u</i></sub> + Source <sub><i>u</i></sub> + Source × Religiosity <sub><i>u</i></sub>	1-to-4.60	0.11

Note. Asterisks mark the preferred model for each hypothesis. The remaining values are the Bayes factors for the respective model versus the preferred model. Subscripts reflect parameter constraints; *u* indicates an unconstrained effect, 1 indicates a common (positive/negative) effect and + indicates a varying positive/negative effect.  $P(\mathcal{M})$  gives the posterior model probability per hypothesis. All models include the covariate level of education.

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<sup>63,65</sup>. Similarly, strong conservatives who distrust science may become less convinced of human-caused global warming when presented with scientific consensus information<sup>62</sup>. In other words, laypeople may apply their own ‘power priors’<sup>66</sup> 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 current 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’<sup>67</sup>) randomly credited to either a spiritual authority or a scientific authority. We assessed: (1) whether trusting scientific experts over spiritual leaders is a general heuristic (that is, the Einstein effect); and (2) to what extent perceivers’ religiosity predicts 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. Whereas religiosity and spirituality are overlapping but not interchangeable constructs<sup>68,69</sup>, self-reported religiosity has been positively associated with belief in spiritual phenomena such as fate, spiritual energy and a connected universe<sup>70–72</sup> (though not unequivocally<sup>73</sup>). Consequently, we expected religiosity to be associated with increased receptivity to gobbledegook from a spiritual authority.

All confirmatory hypotheses and included measures were pre-registered on the Open Science Framework ([osf.io/faj2z/](https://osf.io/faj2z/); the link contains the original preregistration file). The registered component (including additional subprojects) can be found at [osf.io/xg8y5/files](https://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). To further validate the findings from our experimental paradigm, we also analysed a large dataset from

**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 $\times$ 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 (rows 2 and 12). The data indicate: (1) a strong source effect that varies between countries but is consistently positive (scientist > guru), (2) 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.

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<sup>74</sup>.

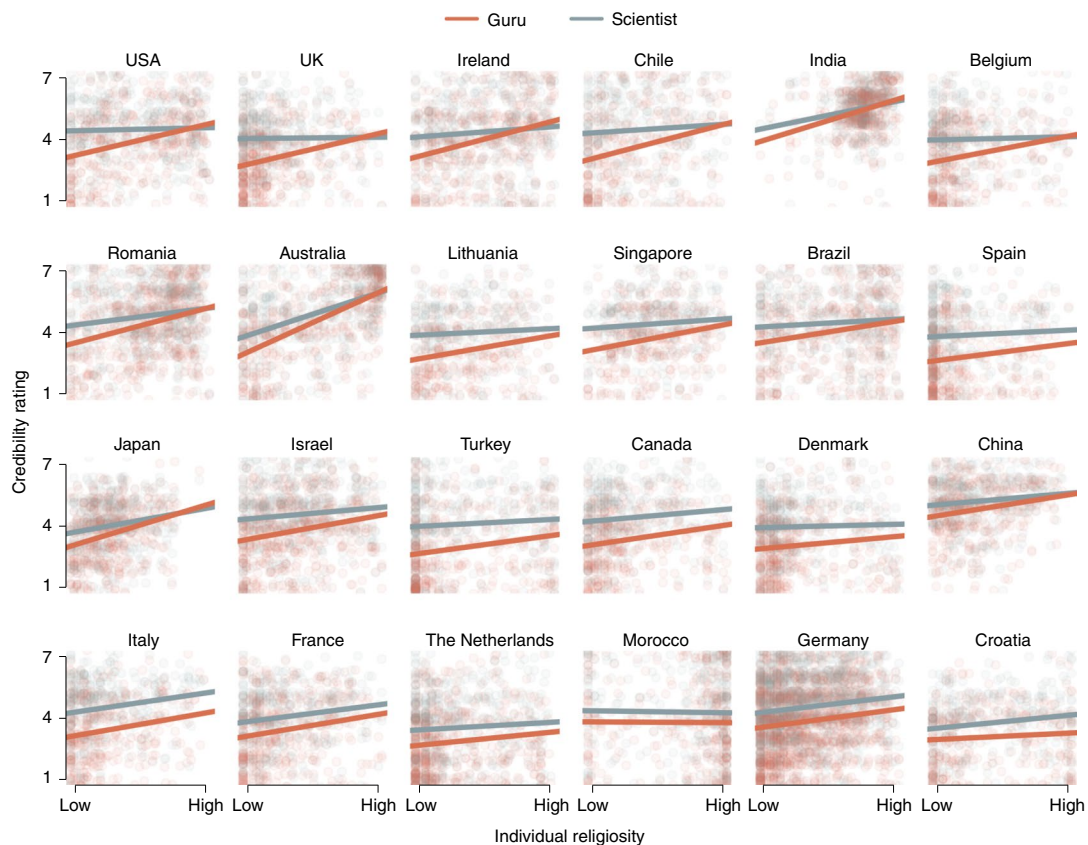
## Results

The two dependent variables that were measured (that is, 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 (95% CI) (0.764, 0.779)) and for the spiritual source (Spearman's  $\rho = 0.827$ , 95% CI (0.822, 0.833); Supplementary Fig. 7)<sup>75</sup>. Because the pattern of results was equal across the dependent variables, we decided to describe only 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 (that is, a scientist versus 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 (Fig. 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 (that is, the scientist and spiritual guru are judged equally credible), the model with a common positive effect of condition across countries (that is, 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 (that is, 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 (that is, 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 most evidence for the positive effects model, which assumes a varying but consistently positive effect across countries. The source effect is favoured  $1.1 \times 10^{210}$ -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 ( $BF_{+1} = 8.9 \times 10^{17}$ ; explained variance (Bayesian  $R^2$ ) is 17.9%, 95% CI (17.0%, 18.7%)). The mean (95% CI) of the unstandardized size of the source effect in the full model is 0.70 (0.60, 0.79) on a seven-point Likert scale and the s.d. between countries is 0.16. Also note that as shown in Fig. 1, the within-country individual differences in credibility ratings are large, indicating that most of the variance is located at the lower level (that is, the individual level). The intraclass correlation coefficients 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 Supplementary Information. 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.

**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 standardized). Our hypothesis states that for individuals with low religiosity, credibility ratings should be higher for gobbledegook from a scientific source than for gobbledegook from a spiritual guru. For highly religious individuals, the reverse effect is expected, that is, higher credibility ratings for gobbledegook ascribed to a guru than for gobbledegook ascribed to a scientist.



**Fig. 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 and circles denote ratings for the spiritual guru and grey lines and circles denote ratings for the scientist. Circles reflect individual observations and are jittered to enhance visibility. Credibility was measured on a seven-point Likert scale.

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,  $BF_{10} = 0.99 \times 10^{15}$  ( $R^2 = 18.1\%$ , 95% CI (17.2%, 19.0%)). The data are uninformative for distinguishing between the common interaction and the varying positive interaction model ( $BF_{1p} = 1.28$ ), indicating that both are equally plausible. Although we cannot conclude whether 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 s.d. between countries is 0.09 on the seven-point Likert scale. As evident from Fig. 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 Fig. 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.

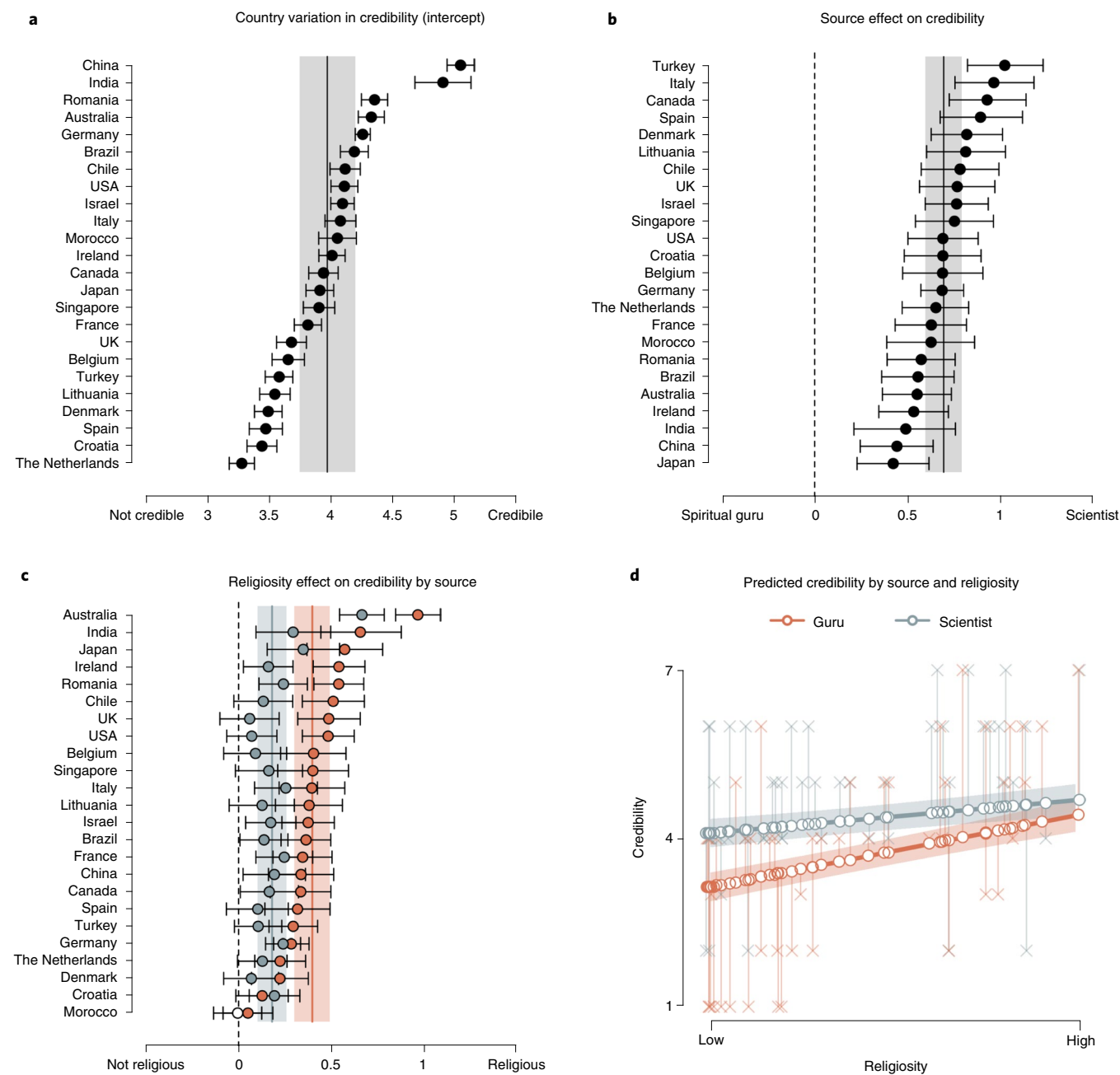
**Exploratory analyses.** In an exploratory fashion, we assessed to what extent the source manipulation influenced the effort participants put into processing the statements. To this end, we looked at: (1) response time for the evaluation of each statement as a proxy for processing time of the message, and (2) memory performance of

words presented in the statements as a proxy for encoding quality. For these exploratory models, we assessed only evidence for a common effect, because visual inspection of the data suggested no or only very small and homogeneous effects (Fig. 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 response time = 28.30 s) than that of the guru (median response time = 27.0 s;  $BF_{10} = 8,050.48$ ). 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 versus the guru ( $BF_{10} = 0.03$ ,  $BF_{01} = 30.78$ ).

**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 five target words (included in the statement) and five 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 anecdotal evidence against a common effect of source on memory performance: participants did not perform better on recognizing words from the statement by the scientist than the statement by the guru ( $BF_{10} = 0.53$ ;  $BF_{01} = 1.90$ ;



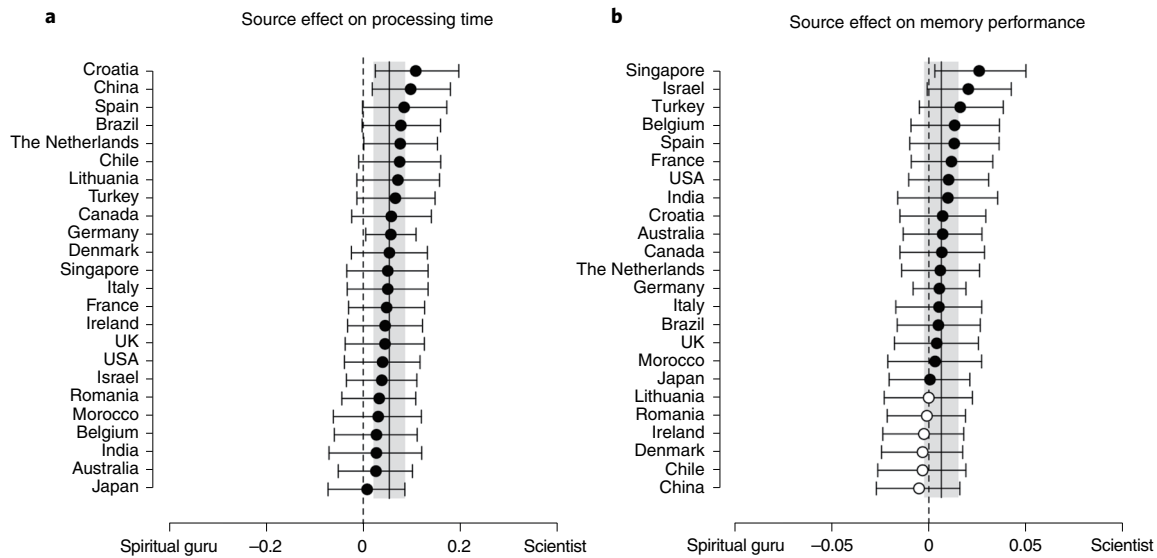
**Fig. 2 | Summary of the multilevel-model (unconstrained) estimates per country and predicted overall effects.** **a,b**, It is apparent that there is substantial variation across the 24 countries in **(a)** overall credibility judgements (that is, intercept) and **(b)** the effect of scientific versus spiritual source. **c**, Individual religiosity has a stronger effect on credibility judgements 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 error bars give the 95% CI for each country. The vertical lines denote the overall estimated effect with the 95% CI in the shaded bands. The dashed lines indicates zero. **d**, Predicted credibility as a function of source and individual religiosity, showing that the difference in credibility ratings for the scientist (grey lines) versus the guru (red lines) is less pronounced for high-religiosity individuals than low-religiosity individuals. The shaded bands reflects the 95% CI, the observed values for two randomly sampled participants per country, and circles reflect the corresponding estimated values. Crosses and circles are jittered to enhance visibility.

standardized estimate = 0.014 (0.001, 0.035)). Finally, there was moderate evidence against an interaction,  $BF_{10} = 0.31$ ,  $BF_{01} = 3.27$ .

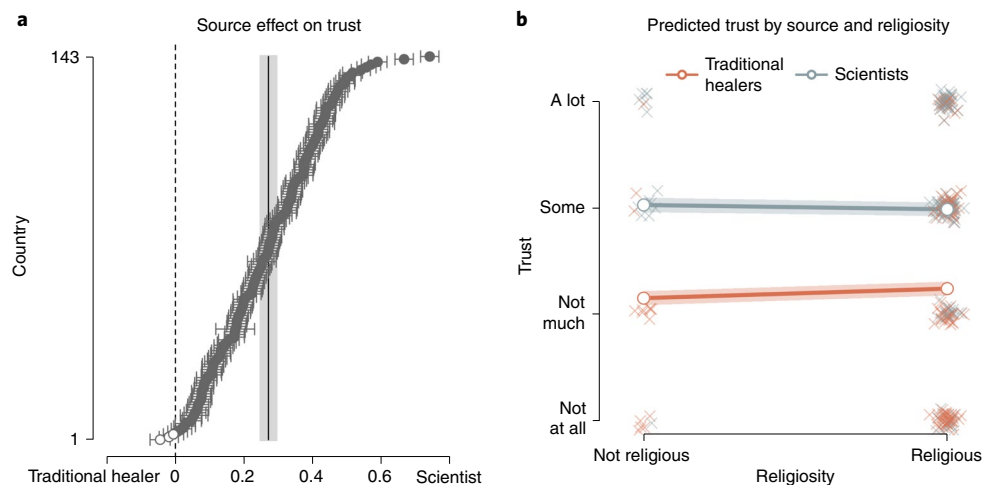
As a sanity check, we showed that there is an extremely strong effect of processing time on memory performance; participants who spent more time processing the statement, also performed better on the memory task ( $BF_{10} = \infty$ ).

**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)). Although the positive effects model strongly outperforms both the null model



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



**Fig. 4 | Multilevel-model (unconstrained) estimates and predicted overall effects for explicit trust ratings. a.** The source effect on trust ratings for each of the 143 countries, showing that in all but three countries, scientists are trusted more than traditional healers. The estimates are ordered from largest to smallest, and open circles denote negatively valued effects. The error bars give the 95% CI for each country. The vertical lines denote the overall estimated effect with the 95% CI in the shaded bands. The dashed lines indicates zero. **b.** 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 with non-religious individuals. The shaded bands reflect the 95% CI, crosses reflect the observed values for two randomly sampled participants per country, and circles reflect the estimated values per condition. The crosses are jittered to enhance visibility.

and the common effect model ( $BF_{+0}$ ,  $BF_{+1} > 10^{308}$ ;  $R^2$  for the positive effects model = 28.1% (27.8%, 28.3%)), the analysis indicates most evidence for the unconstrained model  $\mathcal{M}_{it}$ , which indicates that scientists are not explicitly trusted more than traditional healers in all of the 143 countries ( $BF_{it+} = 320.76$ ). Nonetheless, as displayed in Fig. 4a, only in 3 of the 143 countries is the mean of the estimated source effect negative, whereas the overall effect is clearly positive.

We also investigated the fit effect in this dataset, by including an interaction term between authority (scientists versus traditional healers) and religiosity (religious versus 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,  $BF_{10} = 6.3 \times 10^{14}$ ,  $R^2 = 28.1\%$  (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 versus traditional healers is smaller for religious

individuals than for non-religious individuals. Interestingly, whereas the experimental study found that religiosity was associated with increased credibility ratings for both sources, albeit to a smaller extent for the scientist (Fig. 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 Supplementary Information 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 eight 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 ( $n_{\text{obs}} = 1616$  (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<sup>76</sup>
4. Using the importance rating instead of the credibility rating as the outcome variable
5. Applying a between-subjects design by taking only 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.

The results of these robustness analyses are given in Table 2 and corroborate the conclusions from the main analyses: the data indicate (1) a source effect that varies between countries but is consistently positive (scientist > guru), and (2) a positive source-by-religiosity interaction effect (either a common or varying effect).

## 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 judgements 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 that from the scientist, whereas 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 of 143 of these countries, people indicated greater trust in scientists than in traditional healers, with a larger difference for non-religious compared with religious individuals. Robustness analyses for the experimental study indicated that the effects were robust against different data inclusion criteria (for example, attention checks) and analytic choices (for example, selection of covariates, dependent variable, prior settings). Moreover, the effects also emerged compellingly when analysed as a between-subjects design

(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<sup>77</sup>). Results of exploratory response time analyses suggest that in addition to giving more positive evaluations, people may actually put more effort into processing information from credible sources (although 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 previous 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 the 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 versus horizontally structured societies, general trust in authorities and specific trust patterns toward religious and secular authorities<sup>78–83</sup>. However, although our analysis indicated quantitative differences in the size of the source effect between countries (that is, varying positive effects), we did not find qualitative differences (that is, changes in the direction or presence of the effect). Descriptively, the weakest source effects (that is, 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 faith traditions. However, this explanation remains speculative and we are hesitant to overinterpret the cross-national variability both in the overall credibility judgements and the effect of source. Although 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 (compared with 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, that is, to refrain from passing judgement at all<sup>84–86</sup>. 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 analytical reasoning about the plausibility of a presented claim does not yield any conclusion, the most rational thing to do may be either suspending judgement (selecting the neutral midpoint of the rating scale) or calibrating judgement to previous 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 judgement 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<sup>55</sup>. 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.

**Table 3 | Descriptive statistics per country**

	N	Age (s.d.)	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.

In this study, we intentionally selected authorities that are generally considered benevolent<sup>30,31</sup> 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 previous attitudes (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 versus stars) in addition to the authority's profession. This approach makes it more difficult to single out which specific factor contributes to the source effect (for example, the observed effects might be partly driven by the authority's 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. Because 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. Although this obviously does not prove that 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 (for example, 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<sup>87</sup>.

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<sup>53,88–90</sup>. 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<sup>91–94</sup>. In addition, although trust is probably 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<sup>78,95,96</sup> and receptivity of pseudo-profound and pseudo-scientific nonsense<sup>36,67</sup>. It could thus be that mistrust in science only partially dampens the allure of well-sounding science-related gobbledegook for intuitive reasoners<sup>36</sup>.



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<sup>38,40,97,98</sup>, strong scepticism is normative within the practice of science—as anyone who has experienced peer review will attest. Although it is 150 years since 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<sup>99</sup>. 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<sup>100–103</sup>, the rise of anti-science populist ideologies might diminish or reverse Einstein effects among political partisans. By contrast, individual differences in deference to science<sup>104</sup> may predict enhanced Einstein effects, although a recent study failed to find this pattern for faith in science (van der Miesen et al., in preparation). 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 (for example, the use of jargon, argument coherence, disclosure of uncertainty<sup>105</sup>) and personal attitudes towards the topic<sup>106–108</sup>.

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 (for example, ‘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.

## 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 analytical sample of  $N = 10,195$  participants 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 of 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 (that is, through personal networks), but were

selected to cover six 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 (USA).

**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. Because we applied a Bayesian statistical framework, we needed a minimum of 20 countries to have sufficient data for accurate estimation in cross-country comparisons<sup>109</sup>. 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 as the source effect is within-subjects. Data collection was terminated by 30 November 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 Supplementary Information for details about the project). The full translated survey for each included country can be found at [osf.io/kywjs/](https://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<sup>90</sup>, covering religious behaviours (institutionalized such as church attendance and private such as prayer/mediation), beliefs, identification, values and denomination (see Supplementary Table 5 for the exact items). Besides having high face-validity, these measures have been applied cross-culturally in other studies<sup>79,110,111</sup>. A Bayesian reliability analysis using the *Bayesrel* package<sup>112</sup> 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 versus an opaque dark green chalkboard with physics equations; (2) the accompanying grey-scale photo of the alleged source: a man in robes (photo of José Argüelles) versus a man in an old-fashioned suit (photo of Enrico Fermi); and (3) the reported profession: spiritual leader versus scientist. In addition, in the introductory text, the source was further announced as ‘Saul J. Adrian—a spiritual authority in world religions’ versus ‘Edward K. Leal—a scientific authority in the field of particle physics’, names counterbalanced. 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<sup>67</sup>. 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/qsylvw/](https://osf.io/qsylvw/).

The main outcome variable pertained to judgements of the importance and credibility of gobbledegook, measured on a seven-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 dataset 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. The latter was operationalized as the number of items correctly identified as having appeared in the statement. Participants were presented with a list of ten words, including five targets and five distractors, and were asked to select the words that they recognized 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`<sup>76</sup> to estimate and test the multilevel Bayesian regression models<sup>113,114</sup>. The multilevel Bayesian modelling approach allows us to systematically evaluate the evidence in the data under different models: (1) across all countries the effect is truly null; (2) all countries share a common non-zero effect; (3) countries differ, but all effects are in the same (predicted) direction; and (4) 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 (1) to completely unconstrained (4). We refer to these models as the null model, the common effect model, the positive effects model and the unconstrained model, respectively. Note that although the predictions from model (3) are less constrained than those from model (2), 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 (1) would indicate that people from these 24 countries do not differentially evaluate credibility of claims from a guru or a scientist, evidence for (2) 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 (3) 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 (4) 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<sup>115</sup>, based on the original labels specified by Jeffreys<sup>116</sup>.

For the main effect of source ( $\mathcal{H}_1$ ), we specified the following unconstrained model. Let  $Y_{ijk}$  be the credibility rating for the  $i$ th participant,  $i = 1, \dots, N$ , in the  $j$ th country,  $j = 1, \dots, 24$ , for the  $k$ th condition,  $k = 1, 2$ . Then  $Y_{ijk} \sim N(\mu + \alpha_j + v_i\beta + r_i\delta_j + x_{ik}\gamma_j, \sigma^2)$ . Here, the term  $\mu + \alpha_j$  serves as the baseline credibility intercepts with  $\mu$  being the grand mean and  $\alpha_j$  the  $j$ th country's deviation from the grand mean. The  $\beta$  term reflects the fixed effect of the level of education covariate.  $\delta_j$  is the  $j$ th country's main effect of religiosity on credibility ratings. The crucial parameter here is  $\gamma_j$  which is the source effect for the  $j$ th country. In the common effects model, we will replace  $\gamma_j$  with  $\gamma$ . The variable  $x_{ik} = -0.5, 0.5$  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,  $\sigma^2$  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:  $Y_{ijk} \sim N(\mu + \alpha_j + v_i\beta + r_i\delta_j + x_{ik}\gamma_j + r_i x_{ik}\theta_j, \sigma^2)$ , where  $\theta_j$  is the parameter of interest, the religiosity  $\times$  source interaction effect, with  $r_i x_{ik}$  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).

To systematically investigate which third variables should and should not be included in the statistical model, we used directed acyclic graphs<sup>117</sup> to visually represent the causal relations between the variables in our data<sup>118–120</sup>. 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 directed acyclic graphs created in the R package `ggdag`<sup>121</sup>, both country and level of education were identified as potential confounding factors that warranted inclusion, because they may affect both religiosity<sup>122,123</sup> and overall credibility assessments (for example, due to scepticism). 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, that is, gender<sup>124</sup>, age<sup>125</sup>, socio-economic status<sup>126,127</sup>, 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 Supplementary Figs. 4–6 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<sup>128,129</sup>, 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.<sup>114</sup>, concerning

the scale settings on  $\mu$ ,  $\mu_\theta$  and  $\sigma_\gamma^2, \sigma_\theta^2$ . The scale on  $\mu$ ,  $\mu_\theta$  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  $\sigma_\gamma^2, \sigma_\theta^2$  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 Supplementary Information for details on this analysis).

Second, as a stopping rule, we preregistered that data collection would be terminated (1) when the target of  $n = 400$  per country was reached, or (2) by 30 September 2019. However, due to unforeseen delays in construction of the materials and recruitment, this deadline was extended to 30 November 2019. We did not download or inspect the data until after 30 November.

Third, we preregistered to only include countries where usable data from at least 300 participants was collected (that is, complete data from attentive participants). However, we decided to keep the  $n = 291$  participants from Lithuania in the final sample, because 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`<sup>130</sup> to analyse the data and estimate model parameters. However, we ended up using the `BayesFactor` package<sup>76</sup>. 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 Supplementary Information.

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 (Table 2 and the Supplementary Information).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data collected as part of the experimental study, as well as preprocessed secondary data on explicit trust are provided at <https://doi.org/10.17605/osf.io/qsyvw> (<https://osf.io/qsyvw/>). Raw data on the explicit trust ratings are available at <https://wellcome.org/reports/wellcome-global-monitor/2018>.

## Code availability

Analysis code for all main results and supplementary analyses is available at <https://osf.io/qsyvw/>.

Received: 26 November 2020; Accepted: 6 December 2021;

Published online: 7 February 2022

## References

1. The White House Press Briefing. Remarks by President Trump in Press Briefing (Statements & Releases, 2020); <https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-press-briefing-092420/>
2. Farr, C. Stanford Medical faculty lambaste former colleague and Trump coronavirus advisor Dr. Scott Atlas. (CNBC, 2020); <https://www.cnbc.com/2020/09/10/stanford-medical-faculty-lambaste-former-colleague-and-trump-coronavirus-advisor-dr-scott-atlas.html>
3. Brinol, P. & Petty, R. E. Source factors in persuasion: A self-validation approach. *Eur. Rev. Soc. Psychol.* **20**, 49–96 (2009).
4. Pornpitakpan, C. The persuasiveness of source credibility: A critical review of five decades' evidence. *J. Appl. Soc. Psychol.* **34**, 243–281 (2004).
5. McGinnies, E. & Ward, C. D. Better liked than right: Trustworthiness and expertise as factors in credibility. *Pers. Soc. Psychol. Bull.* **6**, 467–472 (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. *Cogn. Sci.* **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. *Pers. Soc. Psychol. Bull.* **39**, 193–205 (2013).
8. Sperber, D. et al. Epistemic vigilance. *Mind Lang.* **25**, 359–393 (2010).
9. Chaiken, S. & Maheswaran, D. Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *J. Pers. Soc. Psychol.* **66**, 460–473 (1994).
10. Petty, R. E. & Cacioppo, J. T. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* 1–24 (Springer, 1986).

11. Clément, F., Koenig, M. & Harris, P. The ontogenesis of trust. *Mind Lang.* **19**, 360–379 (2004).
12. Jaswal, V. K. & Neely, L. A. Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychol. Sci.* **17**, 757–758 (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 Dev.* **62**, 1334–1351 (1991).
14. Harris, P. L., Koenig, M. A., Corriveau, K. H. & Jaswal, V. K. Cognitive foundations of learning from testimony. *Annu. Rev. Psychol.* **69**, 251–273 (2018).
15. Birch, S. A. J., Akmal, N. & Frampton, K. L. Two-year-olds are vigilant of others' non-verbal cues to credibility. *Dev. Sci.* **13**, 363–369 (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. *Evol. Hum. Behav.* **33**, 46–56 (2012).
17. Mascaro, O. & Sperber, D. The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition* **112**, 367–380 (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 (2015).
19. Fiske, S. T. & Dupree, C. Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proc. Natl Acad. Sci. USA* **111**, 13593–13597 (2014).
20. Bloom, P. & Weisberg, D. S. Childhood origins of adult resistance to science. *Science* **316**, 996–997 (2007).
21. Harris, P. L. *Trusting What You're Told: How Children Learn from Others* (Harvard Univ. Press, 2012).
22. Mercier, H. *Not Born Yesterday: The Science of Who We Trust and What We Believe* (Princeton Univ. Press, 2020).
23. Hahn, U., Harris, A. J. L. & Corner, A. Public reception of climate science: Coherence, reliability, and independence. *Top. Cogn. Sci.* **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. *Evol. Hum. Behav.* **22**, 165–196 (2001).
25. Henrich, J. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton Univ. Press, 2015).
26. Johnson, D. D. & Fowler, J. H. The evolution of overconfidence. *Nature* **477**, 317–320 (2011).
27. Johnson, D. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics* Vol. 2 (Princeton Univ. Press, 2020).
28. Sperber, D. Intuitive and reflective beliefs. *Mind Lang.* **12**, 67–83 (1997).
29. Sperber, D. The Guru effect. *Rev. Philos. Psychol.* **1**, 583–592 (2010).
30. Funk, C. Key Findings About Americans' Confidence in Science and Their Views on Scientists' Role in Society (Pew Research Center, 2020); <https://pewrsr.ch/2Hqg31S>
31. Krause, N. M., Brossard, D., Scheufele, D. A., Xenos, M. A. & Franke, K. Trends—Americans' trust in science and scientists. *Public Opin. Q.* **83**, 817–836 (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 Anal.* **30**, 1520–1538 (2010).
33. Weber, E. U. & Stern, P. C. Public understanding of climate change in the United States. *Am. Psychol.* **66**, 315–328 (2011).
34. McCloskey, M., Washburn, A. & Felch, L. Intuitive physics: The straight-down belief and its origin. *J. Exp. Psychol. Learn. Mem. Cogn.* **9**, 636–649 (1983).
35. Mercier, H. The argumentative theory: Predictions and empirical evidence. *Trends Cogn. Sci.* **20**, 689–700 (2016).
36. Evans, A. M., Sleegers, W. & Mlakar, Ž. Individual differences in receptivity to scientific bullshit. *Judgm. Decis. Mak.* **15**, 401–412 (2020).
37. Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E. & Gray, J. R. The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* **20**, 470–477 (2008).
38. Fernandez-Duque, D., Evans, J., Christian, C. & Hodges, S. D. Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *J. Cogn. Neurosci.* **27**, 926–944 (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 (2008).
40. Eriksson, K. The nonsense math effect. *Judgm. Decis. Mak.* **7**, 746–749 (2012).
41. Gruber, D. & Dickerson, J. A. Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Underst. Sci.* **21**, 938–948 (2012).
42. Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G. & Garry, M. On the (non) persuasive power of a brain image. *Psychon. Bull. Rev.* **20**, 720–725 (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 Dev.* **83**, 779–793 (2012).
44. Van Leeuwen, N. Religious credence is not factual belief. *Cognition* **133**, 698–715 (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. *Psycholog. Relig. Spiritual.* **11**, 339–349 (2019).
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. *J. Pers. Soc. Psychol.* **108**, 515–529 (2015).
48. Liquin, E. G., Metz, S. E. & Lombrozo, T. Science demands explanation, religion tolerates mystery. *Cognition* **204**, 104398 (2020).
49. Mahoney, A. in *The Evolution of Religion: Studies, Theories, and Critiques* (eds Bulbulia, J. et al.) 161–166 (Collins Foundation Press, 2008).
50. Irons, W. in *The Evolution of Religion: Studies, Theories, and Critiques* (eds Bulbulia, J. et al.) 51–57 (Collins Foundation Press, 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 Stud. J.* **42**, 674–697 (2014).
52. Gauchat, G. The cultural authority of science: Public trust and acceptance of organized science. *Public Underst. Sci.* **20**, 751–770 (2011).
53. Gauchat, G. Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *Am. Sociol. Rev.* **77**, 167–187 (2012).
54. Brandt, M. J. & Crawford, J. T. in *Advances in Experimental Social Psychology* (ed. Gawronski, B.) 1–66 (Academic Press, 2020).
55. Levy, N. Due deference to denialism: Explaining ordinary people's rejection of established scientific findings. *Synthese* **196**, 313–327 (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. *Soc. Cogn. Affect. Neurosci.* **6**, 119–127 (2011).
57. McPhetres, J. & Zuckerman, M. Religious people endorse different standards of evidence when evaluating religious versus scientific claims. *Soc. Psychol. Pers. Sci.* **8**, 836–842 (2017).
58. Lobato, E. J. C., Tabatabaiean, S., Fleming, M., Sulzmann, S. & Holbrook, C. Religiosity predicts evidentiary standards. *Soc. Psychol. Pers. Sci.* **11**, 546–551 (2020).
59. Robinson, C. Cross-cutting messages and political tolerance: An experiment using evangelical protestants. *Politi. Behav.* **32**, 495–515 (2010).
60. Munro, G. D. & Ditto, P. H. Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Pers. Soc. Psychol. Bull.* **23**, 636–653 (1997).
61. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science* **185**, 1124–1131 (1974).
62. Cook, J. & Lewandowsky, S. Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Top. Cogn. Sci.* **8**, 160–179 (2016).
63. Jern, A., Chang, K.-m. K. & Kemp, C. Belief polarization is not always irrational. *Psychol. Rev.* **121**, 206–224 (2014).
64. Hahn, U., Harris, A. J. L. & Corner, A. Argument content and argument source: An exploration. *Informal Log.* **29**, 337–367 (2009).
65. Batson, C. D. Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *J. Pers. Soc. Psychol.* **32**, 176–184 (1975).
66. Ibrahim, J. G. & Chen, M.-H. Power prior distributions for regression models. *Stat. Sci.* **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. *Judgm. Decis. Mak.* **10**, 549–563 (2015).
68. Zinnbauer, B. J. et al. Religion and spirituality: Unfuzzifying the fuzzy. *J. Sci. Study Relig.* **36**, 549–564 (1997).
69. Paloutzian, R. F. & Park, C. L. *Handbook of the Psychology of Religion and Spirituality* (Guilford Press, 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. *Int. J. Psychol. Relig.* **29**, 184–203 (2019).
71. McClintock, C. H., Lau, E. & Miller, L. Phenotypic dimensions of spirituality: Implications for mental health in China, India, and the United States. *Front. Psychol.* **7**, 1600 (2016).
72. Wilson, M. S., Bulbulia, J. & Sibley, C. G. Differences and similarities in religious and paranormal beliefs: A typology of distinct faith signatures. *Relig. Brain Behav.* **4**, 104–126 (2013).

73. Rice, T. W. Believe it or not: Religious and other paranormal beliefs in the United States. *J. Sci. Study Relig.* **42**, 95–106 (2003).
74. Wellcome Global Monitor – first wave findings (Wellcome Trust, 2019); <https://wellcome.ac.uk/reports/wellcome-global-monitor/2018>
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  $\rho$ . *J. Appl. Stat.* **47**, 2984–3006 (2020).
76. Morey, R. D. & Rouder, J. N. BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.3 <https://CRAN.R-project.org/package=BayesFactor> (2021).
77. Mummolo, J. & Peterson, E. Demand effects in survey experiments: An empirical assessment. *Am. Politi. Sci. Rev.* **113**, 517–529 (2019).
78. Gervais, W. M. et al. Analytic atheism: A cross-culturally weak and fickle phenomenon? *Judgm. Decis. Mak.* **13**, 268–274 (2018).
79. Stavrova, O. Religion, self-rated health, and mortality: Whether religiosity delays death depends on the cultural context. *Soc. Psychol. Pers. Sci.* **6**, 911–922 (2015).
80. Wave 6 Official Aggregate v. 20150418 (World Values Survey, 2010); <https://worldvaluessurvey.org>
81. Mitkidis, P., Xygalatas, D., Buttrick, N., Porubanova, M. & Lienard, P. The impact of authority on cooperation: A cross-cultural examination of systemic trust. *Adapt. Hum. Behav. Physiol.* **1**, 341–357 (2015).
82. Inglehart, R. Mapping global values. *Comp. Sociol.* **5**, 115–136 (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 Cult. Res.* **29**, 240–275 (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'? *Sociol. Methods Res.* **43**, 15–38 (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? *Int. J. Public Opin. Res.* **12**, 208–216 (2000).
86. Krosnick, J. A. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* **5**, 213–236 (1991).
87. Gligorić, V. & Vilotijević, A. 'Who said it?' How contextual information influences perceived profundity of meaningful quotes and pseudo-profound bullshit. *Appl. Cogn. Psychol.* **34**, 535–542 (2020).
88. O'Brien, T. L. & Noy, S. Cultural authority in comparative context: A multilevel analysis of trust in science and religion. *J. Sci. Study Relig.* **57**, 495–513 (2018).
89. McPhetres, J. & Zuckerman, M. Religiosity predicts negative attitudes towards science and lower levels of science literacy. *PLoS ONE* **13**, e0207125 (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. *J. Exp. Soc. Psychol.* **49**, 1210–1213 (2013).
91. Cacciatore, M. A. et al. Opposing ends of the spectrum: Exploring trust in scientific and religious authorities. *Public Underst. Sci.* **27**, 11–28 (2018).
92. McPhetres, J., Jong, J. & Zuckerman, M. Religious Americans have less positive attitudes toward science, but this does not extend to other cultures. *Soc. Psychol. Pers. Sci.* **12**, 528–536 (2020).
93. Rutjens, B. T. & van der Lee, R. Spiritual skepticism? Heterogeneous science skepticism in the Netherlands. *Public Underst. Sci.* **29**, 335–352 (2020).
94. Rutjens, B. T. Science skepticism across 24 countries. *Soc. Psychol. Pers. Sci.* **13**, 102–117 (2022).
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 (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 (2016).
97. Macdonald, K., Germin, L., Anderson, A., Christodoulou, J. & McGrath, L. M. Dispelling the myth: Training in education or neuroscience decreases but does not eliminate beliefs in neuromyths. *Front. Psychol.* **8**, 1314 (2017).
98. Mayo, R. in *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs* (eds Forgas, J. P. & Baumeister, R. F.) 140 (Routledge, 2019).
99. Peirce, C. S. P. in *The Essential Peirce, Volume 1: Selected Philosophical Writings (1867–1893)*, Vol. 1 (Indiana Univ. 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. *Polit. Psychol.* **33**, 275–299 (2012).
101. Graham, J., Haidt, J. & Nosek, B. A. Liberals and conservatives rely on different sets of moral foundations. *J. Pers. Soc. Psychol.* **96**, 1029–1046 (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. *Pers. Individ. Diff.* **50**, 851–856 (2011).
103. Grafman, J., Cristofori, I., Zhong, W. & Bulbulia, J. The neural basis of religious cognition. *Curr. Dir. Psychol. Sci.* **29**, 126–133 (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 Underst. Sci.* **29**, 800–818 (2020).
105. Corner, A. & Hahn, U. Evaluating science arguments: Evidence, uncertainty, and argument strength. *J. Exp. Psychol. Appl.* **15**, 199–212 (2009).
106. Scurich, N. & Shniderman, A. The selective allure of neuroscientific explanations. *PLoS ONE* **9**, e107529 (2014).
107. Kruglanski, A. W. et al. in *Advances in Experimental Social Psychology*, Vol. 37, 345–392 (Elsevier, 2005).
108. Kahan, D. M., Jenkins-Smith, H. & Braman, D. Cultural cognition of scientific consensus. *J. Risk Res.* **14**, 147–174 (2011).
109. Hox, J. J. C. M., van de Schoot, R. & Matthijsse, S. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Surv. Res. Methods* **6**, 87–93 (2012).
110. 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 (2015).
111. Lun, V. M.-C. & Bond, M. H. Examining the relation of religion and spirituality to subjective well-being across national cultures. *Psychol. Relig. Spiritual.* **5**, 304–315 (2013).
112. Pfadt, J. M. & van den Bergh, D. Bayesrel: Bayesian reliability estimation. R package version 0.7.1 <https://CRAN.R-project.org/package=Bayesrel> (2021).
113. Haaf, J. M. & Rouder, J. N. Developing constraint in Bayesian mixed models. *Psychol. Methods* **22**, 779–798 (2017).
114. Rouder, J. N., Haaf, J. M., Davis-Stober, C. P. & Hilgard, J. Beyond overall effects: A Bayesian approach to finding constraints in meta-analysis. *Psychol. Methods* **24**, 606–621 (2019).
115. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge Univ. Press, 2013).
116. Jeffreys, H. *Theory of Probability* 1st edn (Oxford Univ. Press, 1939).
117. Pearl, J. Causal diagrams for empirical research. *Biometrika* **82**, 669–688 (1995).
118. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Chapman & Hall/CRC, 2016).
119. Pearl, J. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM* **62**, 54–60 (2019).
120. Rohrer, J. M. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Adv. Methods Pract. Psychol. Sci.* **1**, 27–42 (2018).
121. Barrett, M. Ggdag: Analyze and create elegant directed acyclic graphs. R package version 0.2.4 <https://CRAN.R-project.org/package=ggdag> (2021).
122. Albrecht, S. L. & Heaton, T. B. Secularization, higher education, and religiosity. *Rev. Relig. Res.* **26**, 43–58 (1984).
123. Schwadel, P. Does higher education cause religious decline? A longitudinal analysis of the within- and between-person effects of higher education on religiosity. *Sociol. Q.* **57**, 759–786 (2016).
124. Miller, A. S. & Hoffmann, J. P. Risk and religion: An explanation of gender differences in religiosity. *J. Sci. Study Relig.* **34**, 63–75 (1995).
125. Argue, A., Johnson, D. R. & White, L. K. Age and religiosity: Evidence from a three-wave panel analysis. *J. Sci. Study Relig.* **38**, 423–435 (1999).
126. Pyle, R. E. Trends in religious stratification: Have religious group socioeconomic distinctions declined in recent decades? *Sociol. Relig.* **67**, 61–79 (2006).
127. Smith, C. & Faris, R. Socioeconomic inequality in the American religious system: An update and assessment. *J. Sci. Study Relig.* **44**, 95–104 (2005).
128. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *J. Math. Psychol.* **56**, 356–374 (2012).
129. Rouder, J. N. & Morey, R. D. Default Bayes factors for model selection in regression. *Multivar. Behav. Res.* **47**, 877–903 (2012).
130. Bürkner, P.-C. Brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).

## Acknowledgements

This work was supported by funds from the Templeton Foundation (grant number 60663) to M.v.E., the Cogito Foundation (grant number R10917) to R.Mc.K., the Australian Research Council (grant number DP180102384) to N.L. and R.M.R., Templeton Religion Trust (reference TRT0196) to J.A.B., and the French Agence Nationale de la Recherche (reference 17-EURE-0017 FrontCog and 10-IDEX-0001-02 PSL) to S.A. The analysis was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

## Author contributions

M.v.E. and S.H. conceptualized the idea, designed the study and formulated the hypotheses. S.A., T.B., R.B., A.C., C.G., R.G., K.H., C.K., R.Mc.K., A.N., L.Q., A.R., J.E.R., R.M.R., H.T., F.U., R.W., D.X. and S.H. provided cultural knowledge (including

translations) for adjusting the materials to the national context and collected the data. S.H. analysed the data with input from J.A.B. and J.M.H. S.H. wrote the first draft of the manuscript, with major critical input from J.M.H., J.A.B., R.M.R., R.Mc.K. and M.v.E. and additional suggestions from S.A., T.B., R.B., A.C., C.G., R.G., W.M.G., K.H., C.K., N.L., A.N., L.Q., A.R., J.E.R., B.T.R., H.T., F.U., R.W. and D.X.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41562-021-01273-8>.

**Correspondence and requests for materials** should be addressed to Suzanne Hoogeveen.

**Peer review information** *Nature Human Behaviour* thanks Nick Allum, Friederike Hendriks, Jens Koed Madsen and Rens van de Schoot for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Data, analysis code, and stimuli are provided at <https://osf.io/qsyvw/>. Full materials in each language can be found at <https://osf.io/kywjs/>.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Administration of the survey was done via Qualtrics (<https://www.qualtrics.com/>)

Data analysis R and RMarkdown were used to conduct the analyses. The analysis was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data collected as part of the experimental study, as well as preprocessed secondary data on explicit trust are provided at <https://osf.io/qsyvw/>. Raw data on the explicit trust ratings are available at <https://wellcome.org/reports/wellcome-global-monitor/2018downloads-4d1c>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We used a mixed design with one repeated-measures manipulated variable (source) and one individual differences measured variable (religiosity). The data are quantitative.
Research sample	Participants (N=10,195) 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). The average age of the total sample was 33.8 (SD=13.8), with 55.9% women. A breakdown of demographics per country is given in Table 3 in the manuscript. In addition, we used an existing dataset of ratings from 117,191 participants across 143 countries. This dataset is obtained from Gallup's Wellcome Global Monitor First Wave Findings ( <a href="https://wellcome.ac.uk/reports/wellcome-global-monitor/2018">https://wellcome.ac.uk/reports/wellcome-global-monitor/2018</a> )
Sampling strategy	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). 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. 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	All data was collected via the computer. Participants received a link to the Qualtrics survey, either by email, social media or through an online platform. No one could access the data except for the main researcher.
Timing	Data collection started on February 6th, 2019 and was terminated by November 30th, 2019. Ten participants completed the survey after this termination date, as the survey links were kept open (though no longer disseminated).
Data exclusions	In total 10,535 completed the survey. As preregistered, 340 participants (3.23%) were excluded because they failed the attention check, leaving an analytic sample of N = 10,195 from 24 countries.
Non-participation	Of all 11,929 people who started the survey, 1746 (14.6%) did not complete it.
Randomization	Participants were randomly allocated to 1 of 4 experimental conditions, consisting of different text-source matches (text A-spiritual source and text B-scientist or text A-scientist and text B-spiritual source) and source orderings (scientific-spiritual or spiritual-scientific ordered condition).

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement	Material/System
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Dual use research of concern

### Methods

n/a	Involvement	Method
<input checked="" type="checkbox"/>	<input type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/>	MRI-based neuroimaging

## Human research participants

---

Policy information about [studies involving human research participants](#)

Population characteristics

See above.

Recruitment

See above.

Ethics oversight

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).

Note that full information on the approval of the study protocol must also be provided in the manuscript.