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


DOI
10.1038/s41562-021-01273-8

Publication date
2022

Document Version
Final published version

Published in
Nature Human Behaviour

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care)
Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material visible on the library and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
The Einstein effect provides global evidence for scientific source credibility effects and the influence of religiosity

Suzanne Hoogeveen1,2, Julia M. Haaf1, Joseph A. Bulbulia2, Robert M. Ross3, Ryan McKay4, Sacha Altay5, Theiss Bendixen6, Renatas Berniūnas7, Arik Cheshin8, Claudio Gentili1, Raluca Georgescu10, Will M. Gervais11, Kristin Hagel12, Christopher Kavanagh13,14, Neil Levy16, Lin Qiu17, André Rabelo18, Alejandro Neely18, Lin Qu17, André Rabelo18, Jonathan E. Ramsay19, Bastiaan T. Rutjens21 and Michiel van Elk22

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’. A group of infectious disease experts and former colleagues from Stanford, however, publicly criticized Dr Atlas, who is a radiologist, for spreading ‘falsehoods and misrepresentation of science’ through his statements about face masks, social distancing and the safety of community transmission. 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 source1–16. Children and adults are sensitive to the past track record of informants11–16, evidence of their benevolence toward the recipient of testimony17–19, as well as how credible the information is at face value20,21. From an evolutionary perspective, deference to credible authorities such as teachers, doctors and scientists is an adaptive strategy that enables effective cultural learning and knowledge transmission22–25. 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 = mc² 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 Sperber26, in some cases, incomprehensible statements from credible sources may be appreciated not just in spite of, but by virtue of their incomprehensibility, as exemplified by the speech of spiritual or intellectual gurus (the ‘Guru effect’). Here, we investigate to what extent different epistemic authorities affect the perceived value of nonsensical information. To this end, we contrasted judgements of gobbledegook spoken by a spiritual leader with gobbledegook spoken...
by a scientist. In addition, we assessed whether the source effect is predicted by individual religiosity and varies cross-culturally, as a proxy for how scientists and spiritual authorities function as ‘gurus’ for different individuals and within different cultural contexts.

Although source credibility effects have typically been investigated for persuasion in marketing and communication, both science and spirituality may present particularly suitable contexts for inducing strong source effects. Scientists are generally considered competent and benevolent sources\(^{56,59}\) and scientific information is often difficult and counterintuitive\(^{32-34}\). The combination of a credible authority and intangible information can increase the probability of obscure scientific information being accepted, by enhancing perceivers’ reliance on the source\(^{56,59}\). 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\(^{35}\). Some experimental evidence, for instance, suggests that irrelevant neuroscience information\(^{37-39}\) or nonsense mathematical equations\(^{40}\) can boost the perceived quality of presented claims, though note that replication studies suggest that mere brain images may not suffice\(^{41-42}\). Notably, these effects were present only among non-experts (that is, people with little formal neuroscience or mathematical training). This distinction suggests that the appeal of ‘science’ 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\(^{43-45}\). 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\(^{46,47}\). 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\(^{48-51}\). In the absence of the means to rationally evaluate a claim and reliable source information, people poorly 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\(^{52-55}\). In a religious context, Christians were found to be more affected by an intercessory prayer when supposedly performed by a (charismatic) Christian than a non-Christian\(^{56}\) and to require less evidence for religious claims (for example, efficacy of prayer to cure illness) than for scientific claims (for example, efficacy of medication\(^{56,57}\)). 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\(^{58}\). 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\(^{59,60,61}\), 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\(^{62-64}\). By including background beliefs, these Bayesian networks describe how a differential weighing of evidence and even divergent updating (belief polarization) can be considered rational and normative. This may explain, for instance, how strong religious believers can become more convinced of their beliefs in the face of disconfirmatory evidence, especially when their faith is being challenged\(^{65}\). Similarly, strong conservatives who distrust science may become less convinced of human-caused global warming when presented with scientific consensus information\(^{66}\). In other words, laypeople may apply their own ‘power priors’\(^{67}\) 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’\(^{68}\)) randomly credited to either a spiritual 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\(^{66,69}\), self-reported religiosity has been positively associated with belief in spiritual phenomena such as fate, spiritual energy and a connected universe\(^{70,71}\) (though not unequivocally\(^{72}\)). Consequently, we expected religiosity to be associated with increased receptivity to gobbledegook from a spiritual authority.

All confirmatory hypotheses and included measures were preregistered on the Open Science Framework (osf.io/fajz2z/; the link contains the original preregistration file). The registered component (including additional subprojects) can be found at osf.io/x9g8y5/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>
<thead>
<tr>
<th>Table 1</th>
<th>Bayes factor model comparisons to test $H_1$ and $H_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Bayes factor</td>
</tr>
<tr>
<td>$H_1$: Source effect</td>
<td></td>
</tr>
<tr>
<td>$M_{10}$: Country, + Religiosity, $\times$ Source</td>
<td>1 to 10$^{128}$</td>
</tr>
<tr>
<td>$M_{17}$: Country, + Religiosity, + Source</td>
<td>1 to 10$^{77}$</td>
</tr>
<tr>
<td>$M_{7}$: Country, + Religiosity, + Source</td>
<td>*</td>
</tr>
<tr>
<td>$M_{11}$: Country, + Religiosity, + Source</td>
<td>1 to 12.30</td>
</tr>
<tr>
<td>$H_2$: Source-by-religiosity effect</td>
<td></td>
</tr>
<tr>
<td>$M_{13}$: Country, + Religiosity, + Source</td>
<td>1 to 10$^{15}$</td>
</tr>
<tr>
<td>$M_{17}$: Country, + Religiosity, + Source, $\times$ Source</td>
<td>*</td>
</tr>
<tr>
<td>$M_{17}$: Country, + Religiosity, + Source, $\times$ Religiosity</td>
<td>1 to 1.28</td>
</tr>
<tr>
<td>$M_{11}$: Country, + Religiosity, + Source, $\times$ Religiosity</td>
<td>1 to 4.60</td>
</tr>
</tbody>
</table>

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; * indicates an unconstrained effect, 1 indicates a common (positive/negative) effect and + indicates a varying positive/negative effect. $P(M)$ gives the posterior model probability per hypothesis. All models include the covariate level of education.
### Table 2 | Bayes factor of different models for robustness checks

<table>
<thead>
<tr>
<th>Robustness Set</th>
<th>N_{obs}</th>
<th>Estimate (95% CI)</th>
<th>BF_{00}</th>
<th>BF_{10}</th>
<th>Preferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main analysis</td>
<td>20,318</td>
<td>0.70 (0.60, 0.79)</td>
<td>10^{10}</td>
<td>10^{17}</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Excluding source incorrect</td>
<td>18,702</td>
<td>0.78 (0.69, 0.88)</td>
<td>10^{14}</td>
<td>10^{15}</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>Excluding Lithuania (n &lt; 300)</td>
<td>19,736</td>
<td>0.69 (0.59, 0.79)</td>
<td>10^{10}</td>
<td>10^{17}</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>Default prior settings</td>
<td>20,318</td>
<td>0.70 (0.56, 0.84)</td>
<td>10^{10}</td>
<td>10^{15}</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>Importance as outcome variable</td>
<td>20,318</td>
<td>0.53 (0.43, 0.63)</td>
<td>10^{18}</td>
<td>10^{11}</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Between-subjects design</td>
<td>10,159</td>
<td>0.83 (0.68, 0.98)</td>
<td>10^{16}</td>
<td>10^{20}</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Including all subjects</td>
<td>20,980</td>
<td>0.69 (0.59, 0.78)</td>
<td>10^{10}</td>
<td>10^{20}</td>
<td>M_{1}</td>
</tr>
<tr>
<td>No covariates</td>
<td>20,318</td>
<td>0.70 (0.60, 0.79)</td>
<td>10^{19}</td>
<td>10^{17}</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>All covariates</td>
<td>20,318</td>
<td>0.70 (0.60, 0.79)</td>
<td>10^{15}</td>
<td>10^{17}</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>Fit effect (Source x Religiosity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main analysis</td>
<td>20,318</td>
<td>−0.21 (−0.29, −0.14)</td>
<td>10^{15}</td>
<td>0.78</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Excluding source incorrect</td>
<td>18,702</td>
<td>−0.23 (−0.32, −0.15)</td>
<td>10^{17}</td>
<td>4.85</td>
<td>M_{1+}</td>
</tr>
<tr>
<td>Excluding Lithuania (n &lt; 300)</td>
<td>19,736</td>
<td>−0.21 (−0.29, −0.13)</td>
<td>10^{14}</td>
<td>0.90</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Default prior settings</td>
<td>20,318</td>
<td>−0.21 (−0.34, −0.09)</td>
<td>10^{15}</td>
<td>10^{−6}</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Importance as outcome variable</td>
<td>20,318</td>
<td>−0.18 (−0.26, −0.10)</td>
<td>10^{9}</td>
<td>0.02</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Between-subjects design</td>
<td>10,159</td>
<td>−0.22 (−0.33, −0.12)</td>
<td>10^{9}</td>
<td>4.67</td>
<td>M_{1}</td>
</tr>
<tr>
<td>Including all subjects</td>
<td>20,980</td>
<td>−0.22 (−0.29, −0.14)</td>
<td>10^{15}</td>
<td>0.56</td>
<td>M_{1}</td>
</tr>
<tr>
<td>No covariates</td>
<td>20,318</td>
<td>−0.22 (−0.29, −0.14)</td>
<td>10^{14}</td>
<td>0.77</td>
<td>M_{1}</td>
</tr>
<tr>
<td>All covariates</td>
<td>20,318</td>
<td>−0.21 (−0.29, −0.13)</td>
<td>10^{15}</td>
<td>0.09</td>
<td>M_{1}</td>
</tr>
</tbody>
</table>

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.
The interaction term was therefore constrained to be negative, in the sense that the coefficient of the source effect becomes smaller (or negative) with increased religiosity. Note that although the interaction term was constrained to have a negative sign, for consistency, we still refer to the model as the positive effects model.

For hypothesis 2, the model comparison summarized in Table 1 shows that the data provide most evidence for the common source-by-religiosity interaction model, which assumes a consistent interaction effect across countries, $\text{BF}_{10} = 0.99 \times 10^{15}$ ($R^2 = 18.1\%$, 95% CI (17.2%, 19.0%)). The data are uninformative for distinguishing between the common interaction and the varying positive interaction model ($\text{BF}_{1p} = 1.28$), indicating that both are equally plausible. 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; $\text{BF}_{10} = 8,050.48$). 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 ($\text{BF}_{10} = 0.03$, $\text{BF}_{01} = 30.78$).

**Memory performance.** After the rating question, participants were presented with a recall item that required them to indicate which words they recognized from the statement. The list consisted of 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 ($\text{BF}_{10} = 0.53$; $\text{BF}_{01} = 1.90$;
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...
and the common effect model ($BF_{+0} > 10^{308}$, $R^2$ for the positive effects model = 28.1% (27.8%, 28.3%)), the analysis indicates most evidence for the unconstrained model $M_u$, which indicates that scientists are not explicitly trusted more than traditional healers in all of the 143 countries ($BF_{u+} = 320.76$). 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

**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 indicate 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.
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 ($n_{obs} = 1616 (7.95\%)$)
2. Excluding data from Lithuania because $n < 300$ (as preregistered)
3. Using a different, less-informed prior setting for $r$ scale; $r = \frac{1}{22} \approx 0.707$, corresponding to a ‘wide’ prior scale provided in the BayesFactor package\(^a\)
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, gobbledygook from a scientist was considered more credible than the same gobbledygook 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\(^b\)). 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 gobbledygook 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\(^c\)–\(^h\). 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 gobbledygook statements: only a small minority of participants, regardless of their national or religious background, displayed canonical scepticism towards the nonsense statements, and 76% of participants rated the scientist’s gobbledygook 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\(^i\)–\(^m\). 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\(^n\). For instance, the premises of using vaccines (‘inserting a virus prevents disease’) or facts about climate change (‘humans are changing the weather’) are intuitively dubious, yet reputable scientists have convinced many laypeople of their truth.
In this study, we intentionally selected authorities that are generally considered benevolent 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. ‘Argüelles’ (ignoring capitalization or diacritical marks). Only one (Spanish) participant mentioned recognizing both of the sources.

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. 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. 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 and receptivity of pseudo-profound nonsense. It could thus be that mistrust in science only partially damps the allure of well-sounding science-related gobbledegook for intuitive reasoners.

In this study, we intentionally selected authorities that are generally considered benevolent 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. ‘Argüelles’ (ignoring capitalization or diacritical marks). Only one (Spanish) participant mentioned recognizing both of the sources.

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. 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. 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 and receptivity of pseudo-profound nonsense. It could thus be that mistrust in science only partially damps the allure of well-sounding science-related gobbledegook for intuitive reasoners.
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—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 fixed beliefs from the ‘method of science’ in favour of the ‘method of authority’, the role of appeals to scientific authority among scientists remains unclear. 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, the rise of anti-science populist ideologies might diminish or reverse Einstein effects among political partisans. By contrast, individual differences in deference to science 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 investigation. 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) and personal attitudes towards the topic.

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 a final 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 via 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 religious, as well as highly secular societies). Table 3 displays the methods 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 provide sufficient data for accurate estimation (reading and processing time for the statement, as well as depth of processing. The latter was 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. Then they were presented with 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 to estimate and test the multilevel Bayesian regression models. 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 (predictive) direction; (4) each country’s effect is positive if the source is a guru, but when others’ effect is negative the model. 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, while 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 some countries perceive 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, based on the original labels specified by Jeffreys.

For the main effect of source ($\eta_j$), we specified the following unconstrained model. Let $Y_{ij}$ be the credibility rating for the $i$th participant in the $j$th country. $i = 1, \ldots, N$ in the $j$th country, $j = 1, \ldots, 24$, for the $4$th condition, $k = 1, 2$. Then $Y_{ij} = N_{ij} + \theta + \gamma + \delta_j + \epsilon_{ij}$, where $N_{ij}$ is the baseline credibility intercepts with $\theta$ being the grand mean and $\alpha$ the $j$th country’s deviation from the grand mean. The $\beta$ term reflects the fixed effect of the level of education covariate. $\delta$ is the $j$th country’s main effect of religiosity on credibility ratings. The crucial parameter here is $\gamma$ which is the source effect for the $j$th country. In the common effect models, we will replace $\gamma$ with $\gamma_i$. The variable $x_0 = 0.5, 0.5$ if $k = 1, 2$, respectively, where $k = 1$ indicates the scientist and the $k = 2$ indicates the guru condition. The variable $x_1$ is the standardized participant-level education covariate. The variable $x_2$ is the standardized religiosity score for each participant. Finally, $\sigma$ is the variance in credibility ratings across participants.

To test the source-by-religiosity interaction hypothesis for hypothesis 3, the model from (1) is extended by including an interaction term: $Y_{ij} = N_{ij} + \theta_i + \gamma_i + \delta_j + \epsilon_{ij}$, where $\theta_i$ is the parameter of interest, the religiosity $x_1$ source interaction effect, with $\gamma_i$ 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 to visually represent the causal relations between the variables in our data. 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 dagitty, we identified both country and participant-level variables as potential confounding factors that warranted inclusion, because they may affect both religiosity 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, age, socio-economic status, 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, 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., concerning the scale settings on $\mu$, $\rho$ and $\sigma^2$. The scale on $\mu$, $\rho$ reflects the expected size of the overall source effect and source-by-religiosity effect, respectively, and is set to 0.8 (small–medium effect). The scale of $\sigma^2$ reflects the expected amount of variability in these effects and is set to 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, 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 a scientific precision of the estimates. Moreover, it would simply be too unfortunate to remove all data from a highly understudied country.

Fourth, we preregistered that we would use the R package brms to analyse the data and estimate model parameters. However, we ended up using the BayesFactor package. 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/qyvw (https://osf.io/qyvw/). Raw data on the explicit trust ratings are available at https://wellcome.org/reports/wellcome-global-monitor/2018.

Code availability

Code availability for all main results and supplementary analyses is available at https://osf.io/qyvw/.

Received: 26 November 2020; Accepted: 6 December 2021; Published online: 7 February 2022

References

74. Wellcome Global Monitor – first wave findings (Wellcome Trust, 2019); https://publications.wellcomecollection.org/monitor-2019

Acknowledgements
This work was supported by funds from the Templeton Foundation (grant number 60663) to M.v.E., the Cogito Foundation (grant number R190171) to R.M.K., the Australian Research Council (grant number DP180102384) to N.L. and R.M.R., Templeton Religion Trust (reference number TGR(R)10196) to J.A.B., and the French Agence Nationale de la Recherche (reference 17-EURE-0017 FrontCog and 10-IDEIX-0001-02 PSF) to S.A. The analysis was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

Author contributions

**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](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/qsyrw/](https://osf.io/qsyrw/). Full materials in each language can be found at [https://osf.io/kywjs/](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
- [x] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [x] The statistical test(s) used AND whether they are one- or two-sided
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- [x] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [x] 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)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  *Give P values as exact values whenever suitable.*
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [x] 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/qsyw/. 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

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 (https://wellcome.ac.uk/reports/wellcome-global-monitor/2018)

**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 was 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**

<table>
<thead>
<tr>
<th>n/a</th>
<th>Involved in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>☒</td>
<td>Antibodies</td>
</tr>
<tr>
<td>☒</td>
<td>Eukaryotic cell lines</td>
</tr>
<tr>
<td>☒</td>
<td>Palaeontology and archaeology</td>
</tr>
<tr>
<td>☒</td>
<td>Animals and other organisms</td>
</tr>
<tr>
<td>☒</td>
<td>Human research participants</td>
</tr>
<tr>
<td>☒</td>
<td>Clinical data</td>
</tr>
<tr>
<td>☒</td>
<td>Dual use research of concern</td>
</tr>
</tbody>
</table>

**Methods**

<table>
<thead>
<tr>
<th>n/a</th>
<th>Involved in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>☒</td>
<td>ChIP-seq</td>
</tr>
<tr>
<td>☒</td>
<td>Flow cytometry</td>
</tr>
<tr>
<td>☒</td>
<td>MRI-based neuroimaging</td>
</tr>
</tbody>
</table>
# Human research participants

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

<table>
<thead>
<tr>
<th>Population characteristics</th>
<th>See above.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>See above.</td>
</tr>
</tbody>
</table>

| 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.