nature human behaviour

Article

Intermediate levels of scientific knowledge are associated with overconfidence and negative attitudes towards science

Received: 19 September 2022

Accepted: 11 July 2023

Published online: 14 September 2023

Check for updates

Simone Lackner $\mathbb{O}^{1,5}$, Frederico Francisco $\mathbb{O}^{2,5}$, Cristina Mendonça^{1,5}, André Mata³ & Joana Gonçalves-Sá $\mathbb{O}^{1,4}$

Overconfidence is a prevalent problem and it is particularly consequential in its relation with scientific knowledge: being unaware of one's own ignorance can affect behaviours and threaten public policies and health. However, it is not clear how confidence varies with knowledge. Here, we examine four large surveys, spanning 30 years in Europe and the United States and propose a new confidence metric. This metric does not rely on self-reporting or peer comparison, operationalizing (over)confidence as the tendency to give incorrect answers rather than 'don't know' responses to questions on scientific facts. We find a nonlinear relationship between knowledge and confidence, with overconfidence (the confidence gap) peaking at intermediate levels of actual scientific knowledge. These high-confidence/intermediate-knowledge groups also display the least positive attitudes towards science. These results differ from current models and, by identifying specific audiences, can help inform science communication strategies.

It has been argued that 'no problem in judgement and decision making is more prevalent and more potentially catastrophic than overconfidence¹. Overconfidence can be broadly defined as a bias that makes people have a subjective assessment of their own aptitude that is greater than their objective accuracy of such aptitude. These subjective assessments lead to calibration errors with people both overestimating (overconfidence) or underestimating (underconfidence) their ability. A well-studied example concerns how confidence varies with knowledge: does possessing knowledge also come with accurate metacognition about one's knowledge? That is, are the ones who know less aware of it or, conversely, 'Do those who know more also know more about how much they know'²? If metacognition was perfect, one should expect a linear relationship between how much one knows (knowledge) and how much one thinks one knows (confidence) (Fig. 1a, yellow line). However, there is now general agreement that a perfect linear relationship does not exist and that miscalibrations in the internal representation of accuracy can have dire consequences³⁻⁶. Therefore, accurately identifying populations more at risk of overconfidence is fundamental. Notably, ref. 7 have shown that, at best, there is a weak correlation between knowledge and confidence (Fig. 1a, purple dashed line), with the least knowledgeable being more likely to overestimate their skills. In particular, the Dunning–Kruger effect has been identified in controversial antiscience movements, such as vaccine hesitancy and opposition to genetically modified foods^{8,9} and overconfidence can play a role in pandemic control and science communication in general.

Since the publication of the original Dunning–Krueger effect, the field has accumulated a large body of evidence and evolved to suggest a more nuanced pattern than the original proposal. For example, several studies indicate that the main tenet of the effect–that the unskilled are unaware of their lack of skill–might not be universal^{10–13}, such that not all unskilled people are unaware and different authors, including

¹LIP - Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal. ²Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal. ³CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal. ⁴Nova School of Business and Economics, Carcavelos, Portugal. ⁵These authors contributed equally: Simone Lackner, Frederico Francisco, Cristina Mendonça. Se-mail: joanagsa@lip.pt





Dunning, have tried to replicate the effect in different circumstances and countries, finding that overconfidence follows individual and cultural trends¹⁴. Moreover, gauging confidence or knowledge faces its own challenges as the metrics themselves are never perfect and might not be universal or even independent. Therefore, there is reason to believe that the relationship between confidence and knowledge is modulated by several factors¹⁵⁻¹⁷, including methodological as well as analytical issues, among which we highlight two. First, the confidence measures typically used in this research often rely on self-reported metrics, whereby participants are asked how well they believe they performed on a given task and this can introduce important biases: participants often show social desirability bias^{18,19} and struggle to accept limitations publicly^{20,21}, with many people offering answers to survey questions even when the subject is fictitious²². Furthermore, people are often motivated to view themselves in a positive light, which can also lead to (positive) distortions in self-report²³. Moreover, subtler measures of confidence (for example, response times, skin conductance, brain imaging) suggest that the unskilled often show doubt at a more implicit level²⁴. Second, Dunning-Kruger-type studies typically present both knowledge and confidence in comparative scales, in which respondents are asked to compare their performance to the performance of others (for example, student participants in ref. 7 were asked to compare their ability to recognize humour in relation to the average student, using percentile ranking), making it difficult to distinguish between poor self-assessment or underestimation of others' abilities. In addition, these comparative assessments are often presented in quartiles, which can hide important variability, by grouping together respondents with very different mean knowledge or confidence, as Dunning and Kruger have also observed²⁵. However, as the extremes of the distributions often include very little data, not grouping them could also lead to treating noise as signal. Still, it is broadly accepted that the least knowledgeable tend to show the largest confidence gap or, in other words, that the least knowledgeable are often the more overconfident and this was recently restated in relation to controversial science issues, in a study that assesses both objective scientific knowledge and subjective self-assessment²⁶.

Here, we use a different approach to address both these issues. First, we introduce a new metric, based on the premise that 'don't know' versus incorrect answers to knowledge questionnaires can be used as



incorrect (*I*) to 'don't know' (DK) ratios assuming different response models. The *x* axis goes from zero correct answers (all answers are either *I* or DK) to all correct answers (zero *I* and zero DK). Perfect calibration (yellow line, respondents never answer incorrectly); DK-like expectation (green dashed lines, respondents are more likely to answer incorrectly at low knowledge levels); random answering (blue dotted line, respondents are as likely to answer *I* as DK); expected distribution when respondents wage their bets, with incorrect answers growing slowly with *K* (large-dashed grey line).

a proxy for confidence and we examine how this indirect confidence varies with knowledge. Second, we apply our metric to several large surveys, conducted over 30 years in Europe and the United States. Third, we analyse these surveys over their full range of variability (instead of considering only performance quartiles) and compare their results to three different models: the already described perfect-metacognition model expectation that confidence should grow linearly with knowledge, the Dunning–Kruger pattern of almost no relationship between the two variables and a third model that introduces the possibility of respondents guessing, which could lead to an overestimation of individual knowledge.

We find that, unlike what is suggested by previous work, overconfidence is not highest among the least knowledgeable but rather at intermediate-knowledge levels, in all three models. We further test this result using two non-comparative direct confidence metrics and confirm the trend that confidence grows faster than knowledge, leading to populations that have some knowledge but strongly overestimate it. Finally, we investigate public attitudes towards science and find that this intermediate-knowledge group also corresponds to the one displaying the most negative attitudes. We discuss the impact of our findings in the broad field of overconfidence studies and science communication and how our model, if correct, can guide future research.

Results

Nonlinear relationship between knowledge and confidence

Figure 1a depicts possible models for the relationship between knowledge and confidence: in the case of perfect metacognition, confidence should grow linearly with knowledge (yellow line); if confidence is independent from knowledge, we should observe no correlation (horizontal blue dotted line); and if confidence grows faster than knowledge, we could observe a nonlinear relationship between the two variables (example in the solid purple line). In the case of the Dunning–Kruger effect, confidence would either grow moderately with knowledge (dashed purple line) or be largely independent (the 'better than average effect' would also predict a close to horizontal line, with higher than 50% average confidence).

As mentioned, most studies that assess this relationship use different types of tasks but mostly rely on self-reported confidence measures, which may introduce bias and confound the relation between metacognition and motivated processes. Moreover, these results are often presented in quartiles, hiding important variance, particularly in the most extreme groups. To minimize these issues, we introduce an indirect confidence variable (Fig. 1b) and apply it to very large surveys.

The indirect confidence metric is conceptually very simple, relies on widely consensual science knowledge and can be applied to all knowledge questionnaires of the format true/false/don't know: when given those three options, an individual with perfect metacognition should either offer the correct answer or answer 'don't know'. Thus, if incorrect answers correspond to situations where individuals believe they know the answer when in fact they don't, this represents a measure of overconfidence. As the proportion of incorrect answers corresponds to a deviation from the 'ideal' number of zero, it may be interpreted as a calibration error and used to estimate how confidence varies with knowledge. Figure 1b and Extended Data Fig. 1 show some examples: (1) in the case of perfect metacognition, there should be no incorrect answers and no calibration error (yellow line); (2) the Dunning-Krueger effect would predict the fraction of incorrect answers to be higher in lower knowledge groups-dashed-green line); (3) if confidence does not vary with knowledge, the ratio should be similar for all knowledge levels and in the particular case of random answering, both incorrect and 'don't know' answers would have the same probability of occurring (dotted blue line); and (4) if overconfidence grows with knowledge or if respondents guessed only when they did not know the answer, the proportion of incorrect answers should increase (large-dashed grey line). These scenarios give rise to monotonous relationships between confidence and knowledge and, except for the fourth, presume overconfidence to be independent from, or decrease with, knowledge. Therefore, all deviations from these predictions are informative.

To test our metric and eliminate the need for grouping respondents in quartiles we applied it to three large-scale surveys on public understanding of science. These surveys (which we refer to as the Eurobarometer (EB)²⁷, General Social Survey (GSS)²⁸ and Pew Research Center's American Trends Panel (Pew)²⁹ datasets) asked several general science questions in a true/false/don't know format (offering two or more possible options) to a total of 96,039 respondents across the United States and 34 European territories, spanning a 30 year period (Extended Data Table 1 and Supplementary Tables 1 and 2).

For each knowledge level, Fig. 2a,d,g show the average proportion of correct (yellow), incorrect (purple) or 'don't know' (green) answers across participants and Fig. 2b.e.h show the distribution of participants with different proportions of 'don't know' (green) and incorrect answers (purple). Respondents in all three surveys offered a sizable number of incorrect answers and these are not randomly distributed: the fraction of incorrect answers varies with knowledge: as the number of individuals who almost never offered wrong answers shrinks very fast this generates a nonlinear relationship, with respondents on both extremes (very low and very high knowledge) offering proportionally more 'don't know' answers and being more likely to never answer incorrectly (Fig. 2b,e,h, dark green bars). To reduce the risk that 'don't know' answers correspond to low-effort answering, we repeated the analysis after removing all straightliners, defined as participants who offered the exact same answer (true, false or don't know; Supplementary Methods) throughout the knowledge scale and observed a similar pattern (Supplementary Fig. 1).

Overconfidence peaks at intermediate-knowledge levels

Following from our argument that the proportion of incorrect answers can be interpreted as a calibration error, the black dotted line in the middle column of Fig. 2 (marking the fraction of individuals that offered a similar number of incorrect and 'don't know' answers, per knowledge bin; Methods), could be considered a metric of overconfidence. As there are different possible null models regarding the expected proportion of incorrect answers (Fig. 1b and Extended Data Fig. 1) and as varying these assumptions changes the calibration error, we tested different baselines. First, we note that the described models (perfect metacognition, Dunning-Krueger effect) assume not only that different metrics perfectly gauge knowledge but also that the metrics of confidence and knowledge are independent and this is rarely the case (for example, asking respondents to subjectively assess their knowledge before or after they have answered objective knowledge questions might alter their answers³⁰. But this lack of metric independence is particularly obvious for our indirect confidence metric as we are using the same questionnaire to gauge both knowledge and confidence: some respondents can correctly guess the answer when they in fact don't know, appearing more knowledgeable and this might introduce a bias in our results. For example, when there are only three options (as in the EB survey, true/false/don't know) respondents who never guess and always answer 'don't know' when they are unsure, would appear as less knowledgeable and simultaneously less confident. On the other hand, if the number of options in the questionnaire increases (for example, four possible answers and one 'don't know' option, as in the Pew survey), guessing correctly becomes less likely and this effect should be less pronounced. Therefore, we simulated different answering strategies, varying the average knowledge and the proportion of guessers in the population. First, we found the parameter space that best fit the knowledge distributions observed in all three surveys (EB, GSS and Pew; Extended Data Fig. 2 and Supplementary Figs. 2 and 3). The best fits were obtained when ~25% of the simulated respondents were guessers (meaning that 75% of the respondents say 'don't know' when they don't know and the remaining 25% guess) for all surveys. Then, using these fit parameters (proportion of guessers, average population knowledge and variance), we simulated the expected proportion of incorrect answers and used these simulations to estimate our expected values (represented as dashed black lines with arrows in Fig. 2b,e,h,k and detailed in the Methods). Thus, we compared our observed proportion of incorrect answers to four different models: (1) perfect metacognition, (2) random answering, (3) accuracy varying linearly with knowledge and (4) our simulated expectations, from respondents potentially guessing the correct answer. The calibration errors were calculated as the difference between the simulated (or expected) proportion of incorrect answers (null models) and the observed incorrect answers, per knowledge bin. Extended Data Fig. 3 shows the results for all models and the third column of Fig. 2 shows the calibration error using the simulated proportions as the null model. Consistently, the respondents identified as the least knowledgeable are not the most overconfident. In fact, the calibration errors are markedly nonlinear, peaking at intermediate-knowledge levels in the three surveys, despite these surveys having covered different populations and spanning three decades, from the first EB to the Pew survey (a quadratic model fits the data better than linear and constant models for all datasets; see Supplementary Methods and Supplementary Table 3 for statistical details). This is also true for most null models tested and this observation conflicts with both the perfect-metacognition model, which would predict constant, zero calibration error and with the Dunning-Krueger effect, which describes a higher confidence gap (or higher calibration error) at lower knowledge levels.

Nonlinear relation is robust and independent of metric

The observed nonlinear relationship could be specific to our indirect metric, to non-controversial science-related questions and/or to the analysed demographics. To test this, we followed three different approaches and: (1) compared between different demographics; (2) developed a new survey with both our indirect confidence metric and a more traditional direct measure; and (3) applied our metric to a previously and independently published study that reported evidence of the Dunning–Kruger effect in a controversial science-related topic⁹.

To compare across populations, we took advantage of the large span of the EB dataset and repeated the analysis per country, age, gender and educational levels. We found very similar trends across the 34





light green ≥ 0.6 to < 0.8, white ≥ 0.4 to < 0.6, light purple ≥ 0.2 to < 0.4 and dark purple 0 to < 0.2 (respondents with mostly incorrect answers). Dotted black lines show the observed average normalized ratios (respondents are similarly likely to answer incorrectly or don't know) with error bars indicating 99% confidence intervals. Dashed black lines with arrows show the observed average normalized ratio for the simulated agents. Calibration error in **c**, **f**, **i** and **l** is calculated as the difference between the real (observed, dotted black line) and the simulated (dashed black line) curves. DIOC, difference in observed confidence. territories and, as ref. 31 before us, we identified the male middle-aged population as being more likely to never answer 'don't know' (Supplementary Information and Extended Data Fig. 4), also in line with studies from the United States³². However, as the EB does not offer stratified samples, the nonlinear relationship could still be due to the studied demographics. Thus, we developed a new survey (referred to as 'Lackner' throughout the paper). This survey focused on three countries (Portugal, Germany and Norway), with different average education levels that displayed different percentages of 'don't know' avoidance in the EB (red in Extended Data Fig. 4d) and compared different confidence metrics. We started by selecting a sample of respondents equally stratified in terms of age, gender and education and repeated the EB knowledge questionnaire with the true/false/don't know structure (Methods and Supplementary Table 4), Figure 2i shows the proportion of correct (yellow), incorrect (purple) and 'don't know' (green) answers per knowledge level and Fig. 2k shows the distribution of incorrect and 'don't know' answers. As before, the proportion of incorrect answers is smaller at low knowledge levels (Figs. 2l and 3e and Supplementary Fig. 4) and the calibration error, compared to a simulated baseline that replicates the observed knowledge distribution (Supplementary Fig. 3), peaks at intermediate-knowledge levels (Fig. 21). Still, as this analysis focuses on our indirect confidence metric, we also included a direct confidence metric, resembling the format used in several papers that replicate the Dunning-Krueger effect, by asking respondents to estimate the number of items they thought they answered correctly. Figure 3f shows these self-reported answers per actual knowledge bin. We also calculated the calibration error as the difference between the number of self-reported correct answers and the number of actually correctly answered questions (Supplementary Fig. 4). Once again, confidence grows very quickly (and nonlinearly) in the early knowledge bins for both confidence metrics (Fig. 3e, dotted line; 3f, blue columns), for all calibration error measures except for one (Extended Data Fig. 3, bottom row) and in all three countries analysed (Supplementary Fig. 4). Interestingly, the trend in age, education and gender remained, indicating that some demographics might be particularly likely to overestimate their knowledge (Extended Data Fig. 4).

As our study focuses on scientific knowledge (and we do not have access to the original Dunning–Krueger effect data), we used available data from a 2019 paper that included a true/false/don't know knowledge questionnaire (referred to here as the Fernbach survey⁹). This paper used a self-assessed, non-comparative confidence metric and reported evidence of the Dunning–Krueger effect on the controversial topic of genetically modified foods. Again, when we plotted the proportion of incorrect answers (Fig. 3h), we observed that this proportion grows very quickly in the lower knowledge bins so that the least and most knowledgeable remain the more likely to answer 'don't know' instead of incorrectly (dark green bars and dotted line).

Therefore and despite the limitations of the metric, we find that the effect is robust across countries, survey formats, scientific topics, distributions of knowledge and proportions of guessers and that different answering strategies alone cannot explain the observed effect of lower confidence in the lower knowledge bins.

Grouping the lower knowledge levels hides nonlinearity

To our knowledge, this nonlinear relationship between knowledge and confidence has not been described and cannot be predicted by the perfect-metacognition model nor the Dunning–Kruger effect. It is robust and does not require comparative scales. As binning might be a relevant problem, particularly for skewed distributions and for extreme knowledge bins, we suggested that the Dunning–Krueger effect could be hiding important variability, by often relying on quartiles. Again, we tested this in three different ways. First, we plotted the knowledge distributions of all surveys (Fig. 3a,d,g and Extended Data Fig. 2) and confirmed that these are skewed to the right. Second, we repeated the analysis using our confidence metric but representing knowledge and confidence in quartiles (Fig. 3b,e, grey and black lines, respectively). Indeed, by using quartiles, the lower knowledge bins (five bins for the EB, seven for Lackner and four for Fernbach) group together in the first quartile and we now reproduce the Dunning–Krueger effect, with confidence varying little or growing approximately linearly between all knowledge levels, for both the indirect (Fig. 3b,e,h) and direct confidence metrics (Fig. 3f). Importantly, given the large size of the surveys, even the more extreme bins have a sizable number of respondents (>1,000 in the case of the EB; Extended Data Fig. 2), so it is unlikely that this effect is due to randomness caused by low numbers.

Finally, we took advantage of another independent question from the EB survey and used it as a proxy of direct confidence: the EB respondents were asked to estimate how informed they were regarding five different topics, including 'new inventions and technologies' and 'new scientific discoveries' and could answer on a three-point Likert-scale, with 'very well informed', 'moderately well informed' and 'poorly informed' options, plus a 'don't know' option (Methods). There was a positive and significant correlation between answering 'don't know' to the knowledge questions and to the science-related 'how informed' items (Supplementary Fig. 5). We then plotted the fraction of respondents answering that they were poorly (bright yellow), moderately (bright blue) or very well (dark green) informed, against the number of correct knowledge questions answered. These are now two independent metrics, with a self-reported confidence metric and again we observed that the less knowledgeable were more likely to answer that they were poorly informed but this was masked when we plotted the exact same data using quartiles (Extended Data Fig. 5 and Fig. 3c).

In summary, in all studies and for all confidence metrics, representing confidence using the quartile aggregation hides the observation that lower errors are found for low knowledge: the quadratic relation is revealed only when the analysis is done across the full knowledge levels (solid black line versus dotted black lines in Fig. 3b,e,h). It is important to again note that representing the calibration error using either the simulated or the perfect metacognition as the null models penalize the higher knowledge bins, with individuals who answered correctly all but one question appearing as overconfident. In fact, if we were to normalize by knowledge or use either random answering or linear growth of incorrect answers as null models (Extended Data Fig. 3), the largest calibration errors would be found for intermediate-knowledge bins only. Indeed, individuals in the most knowledgeable bins are more likely to offer almost exclusively 'don't know' answers than are individuals in the intermediate-knowledge bins, for all surveys (Fig. 2b,e,h,k, dark green bars).

Altogether, this strongly suggests that, at least in the case of scientific knowledge, possessing some knowledge is more dangerous than having no knowledge, in the sense that the least knowledgeable are in fact aware of their limitations and it is the ones with some knowledge who have less accuracy.

Confidence modulates attitudes towards science

One possible important consequence of overestimating scientific knowledge pertains to public attitudes towards science, as recent studies on controversial science-related topics showed a role of overconfidence on negative attitudes towards science^{8,9}. It is well-established that there is no linear relationship between knowledge and attitudes (unlike what the deficit model predicted^{33,34}) and that such 'attitudes' can vary widely depending on the subject, context, political identity and so on³⁵⁻⁴⁰. Therefore, and since we found the largest confidence gap at intermediate knowledge, we asked whether the least positive attitudes towards science would also be found in those knowledge levels.

Beyond the knowledge and the 'how informed' questions, the EB also included a series of ten independent attitudinal items, with answers that can be grouped in an agree/disagree/don't know format



Fig. 3 | Nonlinear relationship between confidence and knowledge is masked by binning. a–h, Data from EB (a–c), Lackner (d–f) and Fernbach (g,h). a,d,g, Frequency distributions of knowledge for the three datasets, with dashed vertical lines marking the quartiles. b,e,h, Fractions of non-correct answers: dark green represents <0.2 normalized incorrect ratio (respondents with the most incorrect answers), light green \geq 0.2 to <0.4, white \geq 0.4 to <0.6, light purple \geq 0.6 to <0.8 and dark purple \geq 0.8 (respondents with the most incorrect answers). Dotted black line shows average normalized ratio of incorrect answers, with error bars indicating 99% confidence intervals. c, Fraction of respondents reporting to be

(Extended Data Table 2, Supplementary Table 5 and Methods). We first plotted the fraction of 'don't know' answers per knowledge level and found that, for every attitude item, with only small variations, the least knowledgeable are the most likely to offer no opinion to the attitudinal questions (Supplementary Fig. 6). Next, we analysed the attitude dependence on knowledge and found that all relationships are quadratic or asymptotic (the effect was particularly strong in the 'agree' answers, possibly due to the acquiescence bias; Methods). This nonlinear behaviour appears in all items, with what can be argued to be the most negative attitudes appearing at intermediate levels of knowledge (Fig. 4a,c,e and Extended Data Fig. 6), which also correspond to

poorly (yellow), moderately well (blue) or very well (green) informed about new scientific discoveries per knowledge level. **f**, Average confidence (percentage of items self-reported as correct) per knowledge level. In **b**, **c**, **e**, **f** and **h** grey lines with black diamonds show average knowledge ranking per knowledge quartile. Black lines with black squares show average confidence per quartile as average normalized ratios of incorrect to 'don't know' answers in **b**, **e** and **h**, as average fraction of people reporting to be very well or moderately informed in **c**, and as average percentage of items self-reported as correct in **f**.

the highest confidence-to-knowledge ratios (shaded areas in Supplementary Fig. 7).

Therefore, attitudes are neither independent from knowledge, nor do they appear to be more negative in lower knowledge levels, as the deficit model would predict. From our analysis, many of the attitudes that can be identified as negative seem to be modulated by a combination of some knowledge and confidence, with overconfidence appearing at the intermediate-knowledge levels and reflecting itself in more negative attitudes towards science both in controversial and less-controversial topics. This has important implications for science communication and public compliance, discussed below.





Fig. 4 | **Negative attitudes toward science peak at intermediate-knowledge levels.** EB data. **a,c,e**, Stacked bar plots with fractions of agree (orange), neutral (yellow) and disagree (blue) answers for three different attitudinal questions: 'We depend too much on science and not enough on faith' (a), 'Scientific and technological research cannot play an important role in protecting the

environment and repairing it' (c), 'Because of their knowledge, scientific researchers have a power that makes them dangerous' (e). **b**,**d**,**f**, The mean fractions of each answer in **a**, **c** and **e**, respectively, with the shaded standard error of the mean for each respective question in red (agree), yellow (neutral) and blue (disagree).

Discussion

In this paper we propose a new indirect, non-self-reported confidence metric, which does not rely on group comparison and a new methodological approach, which looks at the full knowledge and confidence distributions. Using neutral answers as a proxy for confidence, both for the knowledge and the attitudinal questions, we found that (1) confidence grows much faster than knowledge but (2) this growth is nonlinear, with the largest confidence gaps at intermediate- to high-knowledge levels; (3) this effect is robust across metrics and countries; and (4) the least positive attitudes towards science are found for these high-confidence/intermediate-knowledge groups, in both controversial and non-controversial topics.

This is different from what was previously observed and, although not the first demonstration that a little knowledge can be a dangerous thing (for example, ref. 41), might be due to different reasons, including methodological issues, namely that Dunning–Kruger effect-type studies often rely on quartile or model representations and that previous analysis of EB attitudinal items mostly ignore the so-called neutral answers. In fact, more recent work by Sanchez and Dunning provided evidence of a 'beginners' effect', with confidence being low-to-moderate in the initial stages of learning a new task and growing very quickly after a few rounds of completing it, regardless of actual skill⁴². Despite being a different problem, as there is no correlation between number of trials and scientific knowledge, the fast increase in confidence and slower growth in skill or knowledge also results in nonlinear relationships. Indeed, this effect might be less obvious in questionnaires where there is no rule to be learned and applied to other similar questions (such as deductive-reasoning tasks). Another possibility is that our method uses confidence and knowledge variables that have a similar number of items, whereas Dunning-Krueger effect studies often have more options on the knowledge/aptitude than on the confidence axis (for example, ten different grammar questions versus one single 'How do you compare with our peers on grammar'). Also, whether this effect is particular to scientific or another academic knowledge remains to be tested. Finally, our indirect confidence metric uses the same questionnaire to gauge both confidence and knowledge and this raises important issues regarding metric independence. Our simulations suggest that this alone cannot explain the results but might introduce some bias. Taken as a whole, this study also highlights how all metrics have issues and using different null models and analytical

strategies, including binning, might strongly affect results. It adds to the important discussion of how to measure traits as complex as knowledge or confidence and the value of combining different tools, understanding the limitations of each.

Regarding the variation in attitudes towards science, we observed that the least knowledgeable were also more likely to offer 'neutral' answers to the attitudinal questions, indicating lower confidence (Supplementary Fig. 6). Previous studies^{8,9} described the least positive attitudes in the lowest knowledge bins but this difference might be due to the mentioned methodological differences in data binning and analysis (Supplementary Fig. 7). Still, this effect is much stronger in the EB dataset than in refs. 8,9 and there are at least three (non-mutually exclusive) explanations: (1) the EB surveyed many more people, unveiling important differences in the lower knowledge bins: (2) there is an important time gap between these surveys and the EB dataset, which preceded the wide expansion of the internet and of online social networks (misinformation and polarization have increased⁴³, possibly limiting the quality and diversity of accessible information, effectively creating large groups of misinformed citizens); (3) contrary to refs. 8,9, the EB mostly focuses on non-controversial science issues, while respondents on more contested subjects, such as vaccination, might have access to more information on those subjects (both true and false), have stronger opinions and believe themselves to be right (knowing the scientific consensus despite choosing not to follow it⁴⁴). Moreover, the politicization of science, made obvious during the COVID-19 pandemic, together with an increase in political polarization^{45,46} might deepen this divide. Therefore, if the described EB survey was to be repeated, we might observe differences in the gaps between confidence and knowledge and possibly a stronger polarization in the attitude items. Thus, and despite the known problems of knowledge surveys⁴⁷, a new round of a very similar questionnaire should be considered.

Our results indicate that the least knowledgeable show at least some evidence of good metacognition but that individuals with some knowledge are the most likely to overestimate it and to have less-positive attitudes towards science. Importantly, in all studies, most surveyed individuals are on this intermediate-knowledge and high-confidence level (Fig. 3 and Extended Data Fig. 2): this effect was not relevant in our analysis, as bins were normalized by frequency, but is fundamental at a population level, as those intermediate groups are likely to correspond to a large demographic. Conversely, studying the extremes of the distribution was very difficult using classic surveys (with typically low numbers of respondents) but is increasingly possible in the social media and digital era, allowing for a deeper understanding of diversity and identifying subpopulations.

This has clear implications for current science communication and debiasing approaches and strategies should differ as a function of which model is correct. First, if the Dunning-Krueger effect holds, then interventions should target the most unaware, similar to the deficit model; if the present model holds, then they should target those with some intermediate knowledge (corresponding to most of the population). Second, our model indicates that receptiveness to science will be stronger at the lowest and highest knowledge levels, where the confidence-to-knowledge ratios are also lowest. Therefore, offering information that is incomplete, partial or oversimplified, as science communicators often do, might backfire, as it may offer a false sense of knowledge to the public, leading to overconfidence and less support, further reinforcing the negative cycle. Third, if the lowest support for science comes from the overconfident, these might also be the ones more resistant to new information, especially if it contradicts their certainty, creating a negative reinforcement loop. This resistance can manifest as confirmatory tendencies⁴⁸ or other cognitive biases. One debiasing intervention that has proven effective involves demonstrating to the overconfident that their sense of knowledge is illusory⁴⁹⁻⁵² and it is important to share accurate information, while also conveying humility, both on the scientists' and the lay public's side and without colliding with individuals' values and ideology^{53–56}. Finally, and although our results seem robust across demographics, there were interesting individual differences (for example, between genders; Extended Data Fig. 4), in line with recent work documenting individual variability in overconfidence, which can relate to certain personality dispositions^{57–60}, as well as incautious thinking styles that prompt people to jump to conclusions (that is, to make judgements without sufficient evidence; for example, refs. 61,62) and promote unscientific beliefs such as conspiracy theories^{63,64}. The gender differences (Extended Data Fig. 4) were also in line with previous work revealing differences in self-evaluation^{65,66} and even in controversial science-related topics³².

Taken together, our work suggests that, at least in the case of scientific topics, some knowledge is more dangerous than little knowledge and it is fundamental to develop multidisciplinary approaches, building from psychology, social media and complex systems analysis, to avoid such dangers.

Methods

Experimental design

This study takes advantage of large surveys in public understanding of science (existing or developed by the authors) to study how confidence varies with scientific knowledge. It introduces a new indirect confidence metric and tests it across many countries and years. All computations were performed using R v.4.2.1, Microsoft Excel 16, Wolfram Mathematica 10 and Jupyter Notebook 6.01.

We have complied with all relevant ethical regulations and a Data Protection Impact Assessment was evaluated by a certified Data Protection Officer. The Lackner survey obtained ethical clearance from the Nova School of Business & Economics Scientific Council (where the corresponding author was previously located and where the study started), following independent advice from its installation committee of the ethics review board, reference 13/2020, from 25 March 2020. We have also obtained informed consent from all participants in the Lackner study, who were paid through a third party (Respondi).

Survey datasets

Five different datasets were used, covering a large temporal range in Europe and the United States (Supplementary Table 1). The first three are large-scale surveys conducted by widely recognized entities, which focus on scientific knowledge and attitudes towards science and include scientific knowledge items in a true/false/don't know format or similar. The fourth survey was conducted by us, in 2021, in Germany, Portugal and Norway. The fifth dataset is from a 2019 study on the Dunning–Krueger effect in a controversial science-related topic. For simplicity, we refer to these, respectively, as the EB, GSS, Pew, Lackner and Fernbach throughout the text. The detailed items used for analysis are in Extended Data Table 1 (knowledge items for all databases, except Pew), Supplementary Table 2 (Pew's knowledge items) and Extended Data Table 2 (attitudinal items).

Eurobarometer. The EB dataset was obtained through five rounds of the Eurobarometer Science and Technology campaigns, from 1989 to 2005, surveying 34 territories, including European Union members, candidates at the time and other European Economic Area countries, totalling 84,469 individual interviews (mean age 44.66 yr, range 14–99 yr; 53.21% female)²⁷. Unlike previous and subsequent campaigns, this set tried to gauge both knowledge and attitudes, in a consistent way. As there were differences both in the questions asked and in the possible answers, our dataset results from a harmonization effort that took the November 1992 (EB 38.1) round as a base and identified similar variables in the remaining four rounds²⁷. This harmonized dataset is referred to as the EB dataset throughout the text.

General social survey. The GSS dataset was obtained through seven rounds, biyearly from 2006 to 2018, surveying a panel of adults living in households in the United States (both English- and Spanish-language survey-takers)²⁸. Datasets were homogenized on the basis of knowledge question items, resulting in 7,106 computer-assisted personal interviews (mean age 47.75 yr, range 18–89+ yr; 57.04% female). This dataset is referred to as the GSS dataset throughout the text.

Pew Research Center's American Trends Panel. The Pew dataset was obtained through the Wave 42, conducted by Ipsos Public Affairs (Ipsos) from 7 January to 21 January 2019, surveying a probability-based online panel of adults living in households in the United States, totalling 4,464 online interviews (modal age group 30–49 yr (32.46%), range 18–29 to 65+ yr; 56.00% female)²⁹. This dataset is referred to as the Pew dataset throughout the text.

Our study (Lackner). The Lackner dataset was obtained between April and May 2021 using Respondi (https://www.respondi.com) to recruit a stratified sample of respondents according to gender, age and years of education or age at education completed, covering Portugal, Germany and Norway. The survey was conducted online using the Qualtrics software. We received 1,436 respondents total from which 442 respondents failed the data quality checks (Supplementary Table 4), resulting in 994 respondents total (368 Portugal, 282 Norway and 344 Germany; modal age group 55+ yr (26.16%), range 18–24 to 55+ yr, 52.72% female; 9.19% unfinished questionnaires; Supplementary Methods). This dataset is referred to as the Lackner dataset throughout the text. The questionnaire (in the different languages) and the informed consent forms are annexed in the Supplementary Information.

Study 2 (Fernbach). The Fernbach dataset was obtained from ref. 9 through the Open Science Framework (https://osf.io/t82j3/). The data collection took place in July 2016 using the Qualtrics panel and reached a sample of 1,559 participants from France, Germany and the United States (mean age 48.07 yr, range 17–89 yr; 52.92% female). While the original paper includes other studies, study 2 had the largest sample and all knowledge questions referred to a single scientific knowledge area (genetics) which is relevant to a polarized topic (genetically modified foods). This dataset is referred to as the Fernbach dataset throughout the text.

Knowledge variables

The different datasets have different knowledge items listed in Extended Data Table 1 and Supplementary Table 2. The EB dataset includes 12 questions on general textbook science knowledge in a true/ false/don't know format with the indication 'If you don't know, say so and we will skip to the next'. While the EB dataset includes a thirteenth item about how long the Earth takes to orbit the Sun, we chose to discard it as it was dependent on answering a previous question correctly. The GSS includes nine questions in the exact same wording and true/ false/don't know format as used for EB with the indication 'If you don't know or aren't sure, just tell me so and we will skip to the next question. Remember true, false or don't know'. The Pew dataset includes 11 questions to test knowledge of science facts focusing on life science, earth and other environmental science as well as on applications of scientific principles, such as numeracy and chart reading and the understanding of scientific processes. Answers are collected, differently to the EB and GSS, in a four-option multiple choice format and with the indication 'If you don't know the answer, select not sure'. The Lackner survey included the same 12 questions of the EB and GSS in a true/false/don't know format with the indication 'If you 'don't know', please say so'. Finally, the Fernbach dataset includes ten questions on genetics with a true/false/don't know format that were introduced with the instruction: 'For each of the following statements, please tell us whether you think it is true or false'.

We tested the independence of the knowledge items by calculating Spearman correlations and by performing a principal components analysis using our largest database, EB (Supplementary Figs. 8 and 9 and Supplementary Methods). As for the purposes of this project we were not so much interested in measuring individual knowledge as in finding relations between this measure and the other variables, we created a single knowledge variable, ranging from 0 (no correct answers) to 12 (all questions answered correctly). In the case of zero correct answers, these can have been answered incorrectly or as 'don't know'.

To obtain the knowledge quartiles, we calculated the three values that would result in an approximately equal division of the sample into four groups. As the number of items was reduced (12 in EB to 9 in GSS) and many participants had the same performance, the actual percentage in each quartile was not exactly 25%. As there was no reporting in the literature whether group intervals were typically closed on the left or on the right (for example, are those participants who have a performance equal to Q1 excluded or included in the first group?), we decided to use intervals closed on the left consistently in all databases, as that guaranteed no oversampling of the lowest quartile (observed percentages ranged from 17% to 25%). For Fig. 3, the knowledge quartile y axis positioning (to be compared with the quartile's average confidence as in the Dunning-Kruger effect studies) was calculated through the observed average knowledge ranking of participants in each quartile instead of assuming the theoretical quartile centroid (12.5%, 37.5%, 62.5% and 87.5%).

Confidence in knowledge

Confidence in one's knowledge can be measured in different ways but typically involves asking subjects directly (self-reporting). Common measures include asking subjects how knowledgeable they believe they are regarding specific issues, how well they believe they performed on a certain test (before seeing the actual test results) or asking them to compare (rate or rank) themselves to putative others performing the same test or task. The EB, the Pew and the GSS did not include direct confidence metrics. We propose and use in the paper, an indirect measure of confidence in knowledge, defined as the ratio of incorrect to 'don't know' answers in any knowledge questionnaire, as long as it had the format true/false/don't know (or similar). The rationale is that an incorrect answer corresponds to an overestimation of one's knowledge (more details in the main text). To normalize this ratio, we calculated the proportion of incorrect answers over all non-correct answers, as this allowed the data to be displayed in a way such that 0 would represent a ratio with no incorrect (for example, 6 incorrect and 0 'don't knows'), 1 would represent a ratio with only incorrect (for example, 0 incorrect and 6 'don't knows') and 0.5 would represent a tied ratio (for example, 3 incorrect and 3 'don't knows'), with meaningful in-between numbers (for example, 0.33 would represent a proportion of 1:2 and all equivalent ratios such as 2:4 or 4:8).

To experimentally test the new indirect confidence measure, we compared it with two other more direct measures. First, the EB dataset also included questions about how informed people were in five topics ('I would like you to tell me for each of the following issues in the news if you are very well informed, moderately well informed or poorly informed about it?'): sports news, politics, new medical discoveries, new inventions and technologies, new scientific discoveries (new medical discoveries and new inventions and technologies were not asked in years 2001 and 2002). Participants were provided three response options: 'very well informed', 'moderately well informed' and 'poorly informed', plus a 'don't know' option. Second, the Lackner survey included both the indirect metric and overestimation, which is historically used to measure confidence in psychology and cognitive science. Respondents had to self-report how many of the total items they thought they got correctly ('Of the 12 questions you just answered, how many do you think you answered correctly?').

Simulations

To gauge the impact of guessing on the knowledge distribution (and whether this could explain the observed low confidence at low knowledge levels) different answering strategies were simulated using R v.4.2.1, using ggplot2 for the figures. This was done in four steps.

First, we generated a baseline knowledge distribution to represent the 'true knowledge distribution' of the agents. EB, GSS and Lackner displayed a bell-like distribution and Pew seemed to follow a more linear pattern. Following from this observation we opted to simulate an approximate normal and approximate one-sided triangular distributions. For the normal distribution, this was done by randomly taking 100,000 points, with mean (M) and standard deviation (s.d.) as variable parameters. The parameter space was explored by varying both M and s.d. and starting from the observed parameters in each survey (EB, GSS, Pew, Lackner and Fernbach; Supplementary Fig. 2). As guessing inflates the knowledge distribution, there is little point in exploring higher 'true' knowledge scenarios, so parameter estimation of *M* was asymmetrical, testing mostly scenarios in which true average K was lower than observed. Therefore, M was varied in ten steps of 0.5 (1 above and 9 below), unless it reached zero and s.d. was varied in six steps of 0.5, three above and three below the initial observed s.d., unless it reached zero. In the case of the Pew dataset, which displayed an obviously non-quadratic distribution, a one-side, two-parameter triangular distribution was also tested and we used a regression equation to calculate the frequency of each knowledge bin, with b_0 (intercept) and b_1 (slope) as variable parameters. As before, we started from the observed b_0 and b_1 and tested ten steps for b_0 and six steps for b_1 of 0.005.

Second, we simulated the 'observed knowledge distribution', by adding agents who guess when they don't know the answer and guess correctly with a frequency proportional to the number of options (2% or 50% in the case of the EB, GSS, Lackner and Fernbach and 4% or 25% for Pew). These correct answers were added to that agent's total knowledge score, who would appear to be more knowledgeable than the agents who never guessed. To vary the proportion of guessers in the population, we created 'combined knowledge distributions' by weighing the sum of the two previous distributions. The weight was 0.25, 0.5 or 0.75, corresponding to the percentage of agents who would always guess when did not know the correct answer. The frequency of agents who always answered DK was weighted by 1 minus (–) this parameter.

Third, we estimated the proportion of DK to incorrect answers for each knowledge bin, by finding the proportion of agents that came directly from the DK distribution (of step 1) and the proportion of agents that came from the guessing distribution (of step 2), for the different weighting processes. Thus, the proportion of guessers at a given bin represents the proportion of agents in that bin who would have used a guessing strategy (so that all other answers they provided were incorrect answers) and the proportion of agents who always say DK at a given bin represents the proportion of agents in that bin who would have always said DK when faced with an unknown question (so that all other answers they provided are of the DK type).

Finally, the 'combined knowledge distribution' was compared to the actual knowledge distribution observed in each dataset by calculating the mean squared error (MSE) for each combination of parameters. A heatmap with the different MSE can be found in Supplementary Fig. 2. The parameter and distribution type combination with lowest MSE for each dataset is presented in Supplementary Fig. 3.

In our simulations, we can only reproduce a scenario in which the more overconfident are the least knowledgeable when we simultaneously decrease the mean knowledge of the population to very low values and increase the proportion of guessers to 75% but we also lose the quality of the fit (Supplementary Fig. 3), especially in the extreme bins. As anticipated, this reversion is first observed in the Pew survey, that also has a very non-normal distribution of knowledge levels, something that we cannot explain (Extended Data Fig. 2 and Supplementary Fig. 3).

Calibration error models

To obtain an estimate of overconfidence, different proxies for calibration errors were defined (Extended Data Fig. 3). In the case of the indirect confidence metric the calibration error is calculated as the difference between the observed proportion of incorrect and an expected number of incorrect, according to different null models. For the perfect-metacognition model, the expectation is that there should be no incorrect answers P(I) = 0, varying between 0 (perfect calibration) and 1 (maximum overconfidence). In the case of random answering, the probability of offering the correct, the incorrect or the 'don't know' option should be the same P(I) = P(C) = P(DK) = 1/3 and the error varies between -1 (maximum underconfidence, less incorrect answers than expected) and 1 (maximum overconfidence). For accuracy increasing or decreasing with knowledge, different models can be used. In the simple case of linear increase or decrease, we can use as null models $P(I) = (1 - C)/C_{max}$ or $P(I) = C/C_{max}$, respectively, with $C_{\rm max}$ corresponding to total number of knowledge questions in each questionnaire. In the paper, we use a more natural null, accepting that respondents will use a combination of knowledge and guessing, when answering these questions. Thus, we use the simulations described in the previous section that best fit the observed knowledge distribution in each dataset and the calibration error is the difference between the simulated proportion of incorrect answers for a given knowledge level and the actual observed proportion of incorrect and it varies between maximum overconfidence at 1 (for example, if the simulation expected no incorrect answers and the actual observed average is 100%), perfect calibration at 0 (the number of observed incorrect answers matches the expectation) and maximum underconfidence at 1 (if the simulation expected 100% incorrect answers and the observed average is 0). This allows us to account for the fact that people who use a guessing strategy will have inflated performance in the scientific knowledge scale. Extended Data Fig. 3 shows that in all cases, the lowest errors are found in the lowest knowledge bins.

Attitude variables and confidence in attitudes

The EB dataset contains ten core attitude variables listed in Extended Data Table 2. The number of attitude questions in each EB round varied from year to year but there was a core of ten questions homogenized from all rounds that have been used in subsequent studies. Since the GSS survey only included three of the ten attitude questions in the EB survey (marked with asterisk in Extended Data Table 2) and the Pew survey used different items, the ten EB core questions were selected for analysis and repeated in the Lackner survey. The possible answers to the attitude questions were not consistent over time and were systematized (Supplementary Methods and Supplementary Table 5): in the main text, 'neutral answers' represent the aggregates of 'neither agree nor disagree' and 'don't know' answers. Similarly to what was done for the knowledge questions, the 'agree' and 'disagree' answers to the attitude disagree' answers to the attitude disagree' answers to the attitude disagree' and were compared to the proportion of 'neutral' answers'.

To identify a possible 'attitude' dimension, we expanded the work of ref. 37 and included all Eurobarometers and territories, offering more data and statistical power and the possibility of comparing the results longitudinally. Spearman correlation showed poor correlation between items and the first five components of a principle components analysis represented around 65% of the variances (Supplementary Methods and Supplementary Figs. 8 and 9). Thus, all attitude variables were treated independently. Regardless of the polarity of the questions, 'agree' and 'strongly agree' answers were typically more prevalent than disagreement answers (a common effect, known as 'acquiescence bias'^{36,67}); therefore, the Results focuses on the 'agree' answers, as these tend to show a more obvious effect.

Reporting summary

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

Data availability

Surveys EB, Pew and GSS are publicly available and data and details can be found in refs. 27–29, respectively. The Fernbach study was published in ref. 9 and the authors made the data available. Lackner survey data are available at: https://doi.org/10.5281/zenodo.7920776.

Code availability

All code used for the analysis is available at: https://doi.org/10.5281/ zenodo.7920750.

References

- 1. Plous, S. The Psychology of Judgment and Decision Making (McGraw-Hill, 1993).
- Lichtenstein, S. & Fischhoff, B. Do those who know more also know more about how much they know? Organ. Behav. Hum. Perf. 20, 159–183 (1977).
- Johnson, D. D. P. & Levin, S. A. The tragedy of cognition: psychological biases and environmental inaction. *Curr. Sci.* 97, 1593–1603 (2009).
- 4. Tuchman, B. W. The March of Folly: From Troy to Vietnam (Alfred A. Knopf, 1984).
- 5. Johnson, D. D. P. & Tierney, D. R. The Rubicon theory of war: how the path to conflict reaches the point of no return. *Int. Secur.* **36**, 7–40 (2011).
- 6. Akerlof, G. A. & Shiller, R. J. Animal Spirits: How Human Psychology Drives the Economy and Why it Matters for Global Capitalism (Princeton Univ. Press, 2009).
- 7. Kruger, J. & Dunning, D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**, 1121–1134 (1999).
- Motta, M., Callaghan, T. & Sylvester, S. Knowing less but presuming more: Dunning–Kruger effects and the endorsement of anti-vaccine policy attitudes. Soc. Sci. Med. 211, 274–281 (2018).
- 9. Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y. & Rozin, P. Extreme opponents of genetically modified foods know the least but think they know the most. *Nat. Hum. Behav.* **3**, 251–256 (2019).
- Al-Harthy, I. S., Was, C. A. & Hassan, A. S. Poor performers are poor predictors of performance and they know it: can they improve their prediction accuracy. *J. Glob. Res. Educ. Soc. Sci.* 4, 93–100 (2015).
- Händel, M. & Fritzsche, E. S. Unskilled but subjectively aware: metacognitive monitoring ability and respective awareness in low-performing students. *Mem. Cognit.* 44, 229–241 (2016).
- Miller, T. M. & Geraci, L. Unskilled but aware: reinterpreting overconfidence in low-performing students. J. Exp. Psychol. Learn. 37, 502–506 (2011).
- Urban, M. & Urban, K. Unskilled but aware of it? Cluster analysis of creative metacognition from preschool age to early adulthood. J. Creat. Behav. 55, 937–945 (2021).
- Sanchez, C. & Dunning, D. Cultural patterns explain the worldwide perception/performance paradox in student self-assessments of math and science skill. Soc. Psychol. Pers. Sci. 10, 935–945 (2019).
- McIntosh, R. D., Fowler, E. A., Lyu, T. & Della Sala, S. Wise up: clarifying the role of metacognition in the Dunning–Kruger effect. *J. Exp. Psychol. Gen.* 148, 1882–1897 (2019).
- Gignac, G. E. & Zajenkowski, M. The Dunning–Kruger effect is (mostly) a statistical artefact: valid approaches to testing the hypothesis with individual differences data. *Intelligence* 80, 101449 (2020).
- 17. Jansen, R. A., Rafferty, A. N. & Griffiths, T. L. A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nat. Hum. Behav.* **5**, 756–763 (2021).
- Nederhof, A. J. Methods of coping with social desirability bias: a review. *Eur. J. Soc. Psychol.* 15, 263–280 (1985).

- Larson, R. B. Controlling social desirability bias. Int. J. Mark. Res. 61, 534–547 (2019).
- Cole, J. S. & Gonyea, R. M. Accuracy of self-reported SAT and ACT test scores: implications for research. *Res. High. Educ.* 51, 305–319 (2010).
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41 (2009).
- 22. Bishop, G. F., Tuchfarber, A. J. & Oldendick, R. W. Opinions on fictitious issues: the pressure to answer survey questions. *Public Opin. Quart.* **50**, 240–250 (1986).
- 23. Sedikides, C. & Alicke, M. D. in *The Oxford Handbook of Human Motivation* (ed. Ryan, R. M.) Ch. 17 (Oxford Univ. Press, 2012).
- 24. De Neys, W. Bias and conflict: a case for logical intuitions. Perspect. Psychol. Sci. **7**, 28–38 (2012).
- Williams, E. F., Dunning, D. & Kruger, J. The hobgoblin of consistency: algorithmic judgment strategies underlie inflated self-assessments of performance. *J. Pers. Soc. Psychol.* **104**, 976–994 (2013).
- Light, N., Fernbach, P. M., Rabb, N., Geana, M. V. & Sloman, S. A. Knowledge overconfidence is associated with anti-consensus views on controversial scientific issues. *Sci. Adv.* 8, eabo0038 (2022).
- Bauer, M. W., Shukla, R. & Kakkar, P. Public Understanding of Science in Europe 1989–2005—A Eurobarometer Trend File (GESIS, 2012); https://www.gesis.org/en/eurobarometer-data-service/ search-data-access/eb-trends-trend-files/eb-pus-1989-2005
- Smith, T. W., Davern, M., Freese, J. & Morgan, S. L. General Social Surveys, 1972–2018 (NORC, 2019); https://gss.norc.org/ get-the-data
- Funk, C., Kennedy, B., Johnson, C., Hefferon, M. & Thigpen, C. L. American Trends Panel Wave 42 (Pew Research Center, 2019); https://www.pewresearch.org/science/dataset/ american-trends-panel-wave-42/
- Maki, R. H. in Metacognition in Educational Theory and Practice (eds Hacker, D. J. et al.) Ch. 6 (Routledge, 1998).
- Bauer, M. Socio-demographic correlates of DK-responses in knowledge surveys: self-attributed ignorance of science. Soc. Sci. Inform. 35, 39–68 (1996).
- 32. Hamilton, L. C. Self-assessed understanding of climate change. *Clim. Change* **151**, 349–362 (2018).
- Durant, J. R., Evans, G. A. & Thomas, G. P. The public understanding of science. *Nature* 340, 11–14 (1989).
- 34. Bauer, M. W., Allum, N. & Miller, S. What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understand. Sci.* **16**, 79–95 (2007).
- 35. Wynne, B. Knowledges in context. Sci., Technol. Hum. Values 16, 111–121 (1991).
- Evans, G. & Durant, J. The relationship between knowledge and attitudes in the public understanding of science in Britain. *Public* Understand. Sci. 4, 57–74 (1995).
- Pardo, R. & Calvo, F. Attitudes toward science among the European public: a methodological analysis. *Public Understand. Sci.* 11, 155–195 (2002).
- Hamilton, L. C. Education, politics and opinions about climate change evidence for interaction effects. *Clim. Change* **104**, 231–242 (2011).
- McCright, A. M. Political orientation moderates Americans' beliefs and concern about climate change. *Clim. Change* 104, 243–253 (2011).
- Drummond, C. & Fischhoff, B. Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proc. Natl Acad. Sci. USA* **114**, 9587–9592 (2017).

- 41. Oskamp, S. Overconfidence in case-study judgments. J. Couns. Psychol. **29**, 261–265 (1965).
- Sanchez, C. & Dunning, D. Overconfidence among beginners: is a little learning a dangerous thing? J. Pers. Soc. Psychol. 114, 10–28 (2018).
- 43. Quattrociocchi, W., Scala, A. & Sunstein, C. R. Echo Chambers on Facebook (SSRN, 2016); https://doi.org/10.2139/ssrn.2795110
- 44. Science and Engineering Indicators 2014 Ch. 7 (National Science Foundation, 2014).
- 45. Iyengar, S. & Krupenkin, M. The strengthening of partisan affect. *Polit. Psychol.* **39**, 201–218 (2018).
- 46. Klein, E. Why We're Polarized (Profile Books, 2020).
- Bauer, M. in Handbook of Public Communication of Science and Technology (eds Bucchi, M. & Trench, B.) Ch. 8 (Routledge, 2008).
- Nickerson, R. S. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2, 175–220 (1998).
- Fernbach, P. M., Rogers, T., Fox, C. R. & Sloman, S. A. Political extremism is supported by an illusion of understanding. *Psychol.* Sci. 24, 939–946 (2013).
- Meyers, E. A., Turpin, M. H., Białek, M., Fugelsang, J. A. & Koehler, D. J. Inducing feelings of ignorance makes people more receptive to expert (economist) opinion. *Judgm. Decis. Mak.* 15, 909–925 (2020).
- 51. Rozenblit, L. & Keil, F. The misunderstood limits of folk science: an illusion of explanatory depth. *Cogn. Sci.* **26**, 521–562 (2002).
- Walters, D. J., Fernbach, P. M., Fox, C. R. & Sloman, S. A. Known unknowns: a critical determinant of confidence and calibration. *Manag. Sci.* 63, 4298–4307 (2017).
- Campbell, T. H. & Kay, A. C. Solution aversion: on the relation between ideology and motivated disbelief. J. Pers. Soc. Psychol. 107, 809–824 (2014).
- 54. Feinberg, M. & Willer, R. The moral roots of environmental attitudes. *Psychol. Sci.* **24**, 56–62 (2013).
- Feinberg, M. & Willer, R. From gulf to bridge: when do moral arguments facilitate political influence? Pers. Soc. Psychol. B. 41, 1665–1681 (2015).
- Hornsey, M. J. Why facts are not enough: understanding and managing the motivated rejection of science. *Curr. Dir. Psychol.* Sci. 29, 583–591 (2020).
- 57. Campbell, W. K., Goodie, A. S. & Foster, J. D. Narcissism, confidence and risk attitude. J. Behav. Decis. Mak. **17**, 297–311 (2004).
- Littrell, S., Fugelsang, J. & Risko, E. F. Overconfidently underthinking: narcissism negatively predicts cognitive reflection. *Think. Reason.* 26, 352–380 (2020).
- Macenczak, L. A., Campbell, S., Henley, A. B. & Campbell, W. K. Direct and interactive effects of narcissism and power on overconfidence. *Pers. Indiv. Differ.* **91**, 113–122 (2016).
- Sukenik, S., Reizer, A. & Koslovsky, M. Direct and indirect effects of agreeableness on overconfidence. *J. Individ. Differ.* 39, 174–181 (2018).
- 61. Rodríguez-Ferreiro, J. & Barberia, I. Believers in pseudoscience present lower evidential criteria. *Sci. Rep.* **11**, 24352 (2021).
- 62. Sanchez, C. & Dunning, D. Jumping to conclusions: implications for reasoning errors, false belief, knowledge corruption and impeded learning. *J. Pers. Soc. Psychol.* **120**, 789–815 (2021).
- 63. Kuhn, S. A. K., Lieb, R., Freeman, D. andreou, C. & Zander-Schellenberg, T. Coronavirus conspiracy beliefs in the German-speaking general population: endorsement rates and links to reasoning biases and paranoia. *Psychol. Med.* **52**, 4162–4176 (2022).

- 64. Pytlik, N., Soll, D. & Mehl, S. Thinking preferences and conspiracy belief: intuitive thinking and the jumping to conclusions-bias as a basis for the belief in conspiracy theories. *Front. Psychiatry* **11**, 568942 (2020).
- 65. Beyer, S. Gender differences in the accuracy of self-evaluations of performance. J. Pers. Soc. Psychol. **59**, 960–970 (1990).
- 66. Hill, P. W. H. et al. Science possible selves and the desire to be a scientist: mindsets, gender bias and confidence curing early adolescence. Soc. Sci. **6**, 55 (2017).
- 67. Johns, R. *Likert Items and Scales* Survey Question Bank: Methods Fact Sheet 1 (UK Data Service, 2010).

Acknowledgements

We thank C. Souto-Mayor, M. West and J. Nolasco for initial extraction and analysis of the EB dataset, members of the SPAC group for valuable discussions, Fernbach et al.⁹ for making their survey data available and M. Bauer, T. Paixão, M. Entradas and J. Lobo Antunes for critical reading of the manuscript. We also thank L. Hamilton for independently testing the robustness of our metric and confirming some of our findings. This project was partially funded by Welcome DFRH WIIA 60 2011 and ERC-Starting Grant FARE-853566, both to J.G.S. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.G.S. conceived of this work. All authors contributed to the methodology. S.L., F.F., C.M. and J.G.S were involved in investigation. A.M. and J.G.S. undertook supervision. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-023-01677-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01677-8.

Correspondence and requests for materials should be addressed to Joana Gonçalves-Sá.

Peer review information *Nature Human Behaviour* thanks Ian Brunton-Smith, Lawrence Hamilton and the other, anonymous, reviewer(s) 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.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

 \circledast The Author(s), under exclusive licence to Springer Nature Limited 2023

Article



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | **Model comparison.** Different expectations of the proportions of correct (yellow), incorrect (purple) and 'Don't Know' (green) answers, per knowledge bin (**a**, **c**, **e**, **g**, **i**) or proportion of incorrect (purple) and 'Don't Know' (green) within non-correct answers only, per knowledge bin (**b**, **d**, **f**, **h**, **j**) depending on different expectations of the relationship between confidence and knowledge (**k**). Perfect metacognition (**a**, **b**, yellow solid line in **k**) expects all non-correct answers to be of the 'Don't Know' type. Random answering (**c**, **d**, dotted blue line in **k**) expects a constant and even

proportion of 'Don't Know' and incorrect answers regardless of knowledge bin. If overconfidence decreases with knowledge (**e**,**f**, green lines in **k**), the proportion of incorrect answers should decrease as knowledge increases. If overconfidence increases with knowledge (**i**, **j**, solid purple line in **k**), the proportion of incorrect answers should increase as knowledge increases. If respondents only 'guess' when they do not know the answer, the distribution of incorrect may vary depending on the baseline knowledge and the fraction of incorrect should grow nonlinearly with knowledge (**g**, **h**, large-dash grey line in **k**).



Extended Data Fig. 2 | Knowledge distributions. Knowledge distributions for EB (a), GSS (b), Pew (c), Lackner (d) and Fernbach (e). Absolute frequencies for the first bin in each dataset were: 1179, 107, 165, 2, 42, respectively. Absolute frequencies for the last bin in each dataset were: 2753, 556, 685, 64, 48, respectively.

Simulated Accuracy



Extended Data Fig. 3 | Alternative calibration models. Alternative representation of calibration errors, with different null models. Left axis show the answer proportion with green bars representing observed proportion of 'Don't Know' answers per knowledge bin and purple bars representing observed proportion of incorrect answers per knowledge bin, out of all non-correct answers, for EB (a-d), GSS (e-h), Pew (i-l) and Lackner (m-p). In all plots, solid lines show the expected proportion of incorrect answers (null model) and the dashed line the calibration error calculated as the difference between the observed and the corresponding null. As different null models allow for different



d

0.

-1 and 1. In (a, e, i, m), the null model represents the perfect metacognitive model (yellow lines), in which any incorrect answer represent a calibration error. In (**b**, **f**, **j**, **n**), the null model represents random guessing (blue lines), such that an equal proportion of incorrect and 'Don't Know' answers is expected, regardless of knowledge level. In (c, g, k, o), the null model expects confidence to increase in tandem with knowledge (purple lines). In $(\mathbf{d}, \mathbf{h}, \mathbf{l}, \mathbf{p})$, the null model is the result of the simulations with 25% guessers (dark-grey lines).

High

61.98

53.13

31.25

39.58

34.02

34.94

27.46

67.14

23.96

64.29

54.55

19.70

24.24

42.86

40.00

26.67

40.48

45.24

13.64

37.86

36.90

11.82



Extended Data Fig. 4 | **Demographic analyses for the EB and Lackner surveys.** (**a**-**d**) EB data, (**e**-**g**) Lackner data. (**a**, **e**) Box plot shows the fraction of female (orange) and male (blue) respondents that never say 'Don't know' across (**a**) 31 territories or (**e**) 3 countries. Data was negatively tested for normality using scipy's stats module's normaltest function ($\alpha = 0.001$) and for similarity (twotailed Mann-Whitney U test) in both datasets. Three black asterisks indicate statistical significance with *p* < 0.001 in (**a**) and in (**e**) no significant difference was found. (**b**, **f**) Box plot shows the fraction of different age group bins that never say 'Don't know' across all (**b**) 31 territories or (**f**) 3 Lackner-surveyed countries. Diamond indicates an outlier (values in the panel). A two-tailed Kruskal-Wallis H-test and all pairwise comparisons were found to be significant with post hoc Tukey's tests except for (**b**) 25-39 vs. 40-49 and 40-49 vs. 55+ and no evidence of significance in (**f**) (p = 0.042). (**c**, **g**) Box plot shows the fractions of different bins of age at time of completing their education that never say 'Don't know' across all (**c**) 31 territories or (**g**) 3 Lackner-surveyed countries. Diamonds mark outliers (values in the panel). A two-tailed Kruskal-Wallis H-test and all pairwise comparisons were found to be significant with post hoc Tukey's tests, except for 'Up to 15' vs. 'Still studying' and '16-19' vs. '20 + ', in (**c**) and no evidence of significance in (**g**) (p = 0.036). (**d**) Scatter plot shows for each territory the fraction of respondents that never say 'Don't know' sorted according to latitude of the territory. Black line shows the linear regression with low correlation represented R² = 0.21. (**h**). Table with values for all whiskers (low, 3rd column and high, 7th column) and quartiles (Q1, Median and Q3).

31.82

7.14

14.18

32.58

31.67

8.57

33.33

10.00



Up to 15 16-19

till studying

G



Extended Data Fig. 5 | Answer distributions to the 'How Informed' questions and calibration errors. (a, c, e, g, i) Stacked bar plots showing fraction of respondents who answered 'Poorly' (yellow), 'Moderately well' (light blue) and 'Very well' (dark green) when questioned how informed they were about (a) new inventions and technologies, (c) new medical discoveries, (e) new scientific discoveries, (g) politics and (i) sports news, per knowledge level. In all panels, black solid lines with squares indicate mean fraction of respondents who answered 'Moderately well' or 'Very well' per quartile, while solid grey line shows average knowledge rank per quartile. (**b**, **d**, **f**, **h**, **j**) Plot showing the difference between average fraction of respondents who answered 'Moderately well' or 'Very well' per quartile and average knowledge rank per quartile, each represented by a circle marking the average and a vertical line marking the variation in average between bins of the same quartile.



Extended Data Fig. 6 | **EB attitudinal data.** (**a**, **c**, **e**, **g**, **i**, **k**, **m**) show stacked bar plots with fractions of Agree (orange), Neutral (yellow) and Disagree (red) answers in response to 7 EB attitude questions. Order of stacked bars is inverted in (**e**, **k**) as, in those two items, a negative attitude could be revealed by the Agree answer, while the reverse might be true for (**a**, **g**, **m**). (**c**) and (**i**) show a more nuanced response. Figures in (**b**, **d**, **f**, **h**, **j**, **l**, **n**) show the mean fractions across 34 EU territories with standard error of the mean.

Extended Data Table 1 | Knowledge questions

Question	Answer options	Surveys
"The centre of the Earth is very hot."	"True", "False"	1, 2, 4
"The oxygen we breathe comes from plants."	"True", "False"	1, 4
"Radioactive milk can be made safe by boiling it."	"True", " False "	1, 4
"Electrons are smaller than atoms."	" True ", "False"	1, 2, 4
"The continents on which we live have been moving	" True ", "False"	1, 2, 4
their location for millions of years and will continue to		
move in the future."		
"It is the father's gene which decides whether the baby	" True ", "False"	1, 2, 4**
is a boy or a girl."		
"The earliest humans lived at the same time as the	"True", " False "	1, 4
dinosaurs."		
"Antibiotics kill viruses as well as bacteria."	"True", " False "	1, 2, 4
"Lasers work by focusing sound waves."	"True", " False "	1, 2, 4
"All radioactivity is man -made."	"True", " False "	1, 2, 4
"Human beings, as we know them today, developed	" True ", "False"	1, 2, 4
from earlier species of animals."		
"Does the earth go around the sun or does the sun go	"The sun goes around the earth", " The	1, 2, 4
around the earth?"	earth goes around the sun"	
"Yeast for brewing beer or making wine consists of	" True ", "False"	5
living organisms."		
"Ordinary tomatoes do not contain genes, while	"True", " False "	5
genetically modified tomatoes do."		
"The cloning of living things produces genetically	" True ", "False"	5
identical copies."		
"By eating a genetically modified fruit, a person's genes	"True", "False"	5
could also become modified."		
"It is possible to find out in the first few months of	" True ", "False"	5
pregnancy whether a child will have Down's		
Syndrome."		
"Genetically modified animals are always bigger than	"True", " False "	5
ordinary ones."		
"More than half of human genes are identical to those	" True ", "False"	5
of a chimpanzee."		
"It is not possible to transfer animal genes into plants."	"True", " False "	5
"Human cells and human genes function differently	"True", " False "	5
from those in animals and plants."		
"Embryonic stem cells have the potential to develop	" True ", "False"	5
into normal humans."		

List of science knowledge questions used for data analysis from the 4 surveys that only offered True/False/Don't Know format*. each of the five surveys. Eurobarometer (1), General Social Survey (2), Pew (3)*, Lackner (4) and Fernbach (5). * The Pew survey offered several options and details can be found in Supplementary Table 2 ST2, in the supplementary materials. ** Note: From Eurobarometer 2005 onward the word 'father' was replaced with 'mother' in this question and the correct answer became 'False'. The Lackner survey used the 'mother's gene' formulation.

Extended Data Table 2 | Attitude questions

Question

*"Science &	Technology are	making our li	ves healthier	easier and more	comfortable "
Science a	reennology are	making our n	ves neuriner,	cusici una more	connortable.

+"Thanks to scientific and technological advances, the earth's natural resources

will be inexhaustible."

"We depend too much on science and not enough on faith."

⁺"Scientific and technological research cannot play an important role in protecting the environment and repairing it."

+"Scientists should be allowed to do research that causes pain and injury to animals like dogs and chimpanzees if it can produce information about human health problems."

*"Because of their knowledge, scientific researchers have a power that makes them dangerous."

+"The application of science and new technology will make work more interesting."

+"For me, in my daily life, it is not important to know about science."

*"Science makes our way of life change too fast."

** "Thanks to science and technology, there will be more opportunities for the future generations."

List of attitudes towards science questions used for data analysis from the Eurobarometer (EB) dataset. For each statement respondents were asked to state their agreement or disagreement. Starred items (*) indicate items present in General Social Survey. Items marked with a dagger (†) were not part of EB in 1989.

nature portfolio

Corresponding author(s): Gonçalves-Sá

Last updated by author(s): May 16, 2023

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	firmed
		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
\boxtimes		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.
		A description of all covariates tested
		A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	\boxtimes	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.
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
\boxtimes		For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
		Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated
		Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

Software and code

 Policy information about availability of computer code

 Data collection
 For the Lackner survey, the only unpublished dataset in the manuscript, data was collected using an on-line Qualtrics (April 2021 version) survey applied via a third party (Respondi)

 Data analysis
 All computations were performed using R 4.2.1, Microsoft Excel 16, Wolfram Mathematica 10 and Jupyter Notebook 6.01. All code used in simulations is available in a public repository and can be accessed here: https://doi.org/10.5281/zenodo.7920750

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 Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Surveys EB, Pew and GSS are publicly available and data and details can be found in [manuscript references [26], [27] and [28], respectively. The Fernbach study was

published in [9] and the authors made the data available. Lackner survey data is available in a public repository and can be accessed here: https://doi.org/10.5281/ zenodo.7920776

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The publicly available datasets (from Eurobarometer, Pew and GSS) have individual-level surveys classified by self-reported gender. The survey developed by the authors (Lackner) used a stratified sample of respondents that included 524 females and 470 males (self-reported) and the results broadly apply to both genders. Differences between genders may exist and were discussed in the manuscript for the Eurobarometer data (as it represents the largest sample) and for the Lackner survey.
Population characteristics	The study includes close to 90.000 interviews from over 30 European countries and the USA, from a broad number of backgrounds, education levels and ages, with binary self-reported genders represented. The study includes 1) in the EB dataset, 84469 individual interviews (Mean age = 44.66, range 14 to 99; 53.21% female) in five rounds from 1989 to 2005, surveying 34 territories, including EU members, candidates at the time, and other European Economic Area (EEA) countries; 2) in the GSS dataset, 7106 computer-assisted personal interviews (Mean age = 47.75, range 18 to 89+; 57.04% female) in seven rounds from 2006 to 2018, surveying a panel of adults living in households in the United States; 3) in the Pew dataset, a single round in 2019, surveying a probability-based on-line panel of adults living in households in the United States, totaling 4464 on-line interviews (Modal age group = 30-49 (32.46%), range 18-29 to 65+; 56.00% female); 4) in the Lackner dataset, a single survey in 2021 of a stratified sample of respondents according to gender, age, and years of education or age at education completed, covering Portugal, Germany, and Norway, 994 respondents total (368 Portugal, 282 Norway, 344 Germany; Modal age group = 55+ (26.16%), range 18-24 to 55+, 52.72% female); and 5) in the Fernbach dataset, a single survey in 2016 with a sample of 1559 participants from France, Germany, and USA (Mean age = 48.07, range 17-89; 52.92% female). Demographic analyses are also discussed in the manuscript.
Recruitment	For the publicly available surveys recruitment was described elsewhere. The Lackner dataset was obtained between April and May 2021 using Respondi (https://www.respondi.com) to recruit a stratified sample of respondents according to gender, age, and years of education or age at education completed, covering Portugal, Germany, and Norway. For proper comparison between countries in terms of education and age, we requested a sample of 480 respondents for each country DE, NO and PT, with equal quotas for gender, age and years of education / age at education completed. Due to the exclusion of the number of respondents failing quality checks and the difficulty of the recruitment company (Respondi) to recruit more panelists to fill the gaps, data collection was stopped and the final sample is close to representative of each country including self-reported sex (close to 50% each) and educational levels. All participants were volunteers, recruited by a third party company (through which all communication took place, i.e., the researchers never met or contacted the participants) and that did not know the goals of the study. Therefore, it is unlikely that participants self-selected themselves for this particular study. It is not impossible that they had encountered some of the questions in the Lackner survey, but we found no qualitative difference between the Lackner and the other described surveys.
Ethics oversight	We have complied with all relevant ethical regulations and a Data Protection Impact Assessment was evaluated by a certified DPO. The Lackner survey obtained ethical clearance from the Scientific Council of Nova School of Business and Economics - Universidade Nova de Lisboa (where the corresponding author was previously located and where the study started), following independent advice from its installation committee of the ethics review board (CICE), reference 13/2020, from 25/03/2020

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

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 <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Behavioural & social sciences study design

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

Study description	The study aimed at clarifying the relationship(s) between scientific knowledge and confidence and attitudes towards science. It is correlational and quantitative, analysing close to 90.000 individual surveys. Five different datasets were used, covering a large temporal range in Europe and the USA. The first three are large-scale surveys conducted by widely recognized entities, that focus on scientific knowledge and attitudes towards science and include scientific knowledge items in a True/False/Don't Know format or similar. The fourth survey was conducted by us, in 2021, in Germany, Portugal, and Norway. The fifth dataset is from a 2019 study on the Dunning-Krueger effect in a controversial science-related topic.
Research sample	The study includes close to 90.000 interviews, covering more than 30 years and over 30 European countries and the USA, from a

Research sample	broad number of backgrounds, education levels and ages, with both binary genders almost equally represented. Five independent surveys were analysed: Eurobarometer, Pew, GSS, Lackner (this study) and Fernbach. These studies were selected for including knowledge questionnaires in the True/False/Don't know format, or similar. The EB also included a set of questions regarding public attitudes towards science and another that asked participants to declare how informed they were regarding several different topics. As for the samples: 1) in the EB dataset, 84469 individual interviews (Mean age = 44.66, range 14 to 99; 53.21% female) in five rounds from 1989 to 2005, surveying 34 territories, including EU members, candidates at the time, and other European Economic Area (EEA) countries; 2) in the GSS dataset, 7106 computer-assisted personal interviews (Mean age = 47.75, range 18 to 89+; 57.04% female) in seven rounds from 2006 to 2018, surveying a panel of adults living in households in the United States; 3) in the Pew dataset, a single round in 2019, surveying a probability-based online panel of adults living in households in the United States, totalling 4464 online interviews (Modal age group = 30-49 (32.46%), range 18-29 to 65+; 56.00% female); 4) in the Lackner dataset, a single survey in 2021 of a stratified sample of respondents according to gender, age, and years of education or age at education completed, covering Portugal, Germany, and Norway, 994 respondents total (368 Portugal, 282 Norway, 344 Germany; Modal age group = 55+ (26.16%), range 18-24 to 55+, 52.72% female); and 5) in the Fernbach dataset, a single survey in 2016 with a sample of 1559 participants from France, Germany, and USA (Mean age = 48.07, range 17-89; 52.92% female). This summary mostly focuses on Lackner, the only unpublished survey in the manuscript.
Sampling strategy	The researchers had no participation in sample selection for the the Eurobarometer, Pew, GSS, and Fernback studies. The EB, Pew and GSS used random selection, Fernbach used quota matched sampling through a Qualtrics panel. For the Lackner study, one main goal was to control for demographic differences, observed in the EB study, and thus used a quota matched sampling from Respondi so that all age bins (same bins as defined in the EB were similar in terms of educational levels and gender in Germany, Portugal and Norway. These countries were chosen for displaying different behaviours in the EB study. The sample limits were imposed by Respondi.
Data collection	The researchers had no participation in sample selection and interviewing for the the Eurobarometer, Pew, GSS, and Fernback studies. EB used phone interviews, GSS used face to face or (less common) phone interviews, Pew, Lackner and Fernbach respondents answered on-line surveys. In the case of the Lackner study, participants were recruited through a third party company, Respondi, answered an on-line survey, and researchers had no role in selection/interviewing. Moreover, the Respondi recruiters did not know the goal of the study and cannot have influenced selection.
Timing	Data collections for EB, Pew, GSS and Ferbach were discussed in the original reports or papers. For Lackner, data was collected between April and May 2021
Data exclusions	Exclusions for EB, Pew, GSS and Ferbach were discussed in the original reports or papers. For Lackner, we used the following criteria for exclusion: *1) We dropped all respondents that failed the first attention check in the first third of the survey that explicitly calls out for paying attention. (2) We dropped all respondents that failed two or more of the remaining three attention checks (list in the manuscript). (3) The survey was expected to take 15 min reading time (250 words per min). We dropped all respondents that were faster than half that time (7.5min). The total of excluded respondents was 422.
Non-participation	Non-participation rate (i.e., percentage of people who did not take part in the study when offered the chance to do so) for the Lackner dataset was not communicated to us by the recruiting company, but 91% of participants who started the study completed their participation. Non-participation was not reported in Fernbach, was of 23% in Pew, and varied between years in GSS and years and countries in EB (few details provided in existing code books).
Randomization	Randomization was not used as we only had control over the methodology of the Lackner dataset and, in it, we were aiming to replicate the results of the correlational, non-randomized Eurobarometer dataset. Moreover, we were not testing different conditions nor instruments and all participants answered the same questions in the same order.

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

Methods

n/a	Involved in the study
\boxtimes	Antibodies
\boxtimes	Eukaryotic cell lines
\boxtimes	Palaeontology and archaeology
\boxtimes	Animals and other organisms
\boxtimes	Clinical data
\boxtimes	Dual use research of concern

n/a	Involved in the study
\ge	ChIP-seq
\boxtimes	Flow cytometry

MRI-based neuroimaging