



# A strategy to improve expert technology forecasts

Tamara Savage<sup>a</sup>, Alex Davis<sup>a</sup>, Baruch Fischhoff<sup>a</sup>, and M. Granger Morgan<sup>a,1</sup>

Edited by B. L. Turner, Arizona State University, Tempe, AZ, and approved March 4, 2021 (received for review January 15, 2021)

**Forecasts of the future cost and performance of technologies are often used to support decision-making. However, retrospective reviews find that many forecasts made by experts are not very accurate and are often seriously overconfident, with realized values too frequently falling outside of forecasted ranges. Here, we outline a hybrid approach to expert elicitation that we believe might improve forecasts of future technologies. The proposed approach iteratively combines the judgments of technical domain experts with those of experts who are knowledgeable about broader issues of technology adoption and public policy. We motivate the approach with results from a pilot study designed to help forecasters think systematically about factors beyond the technology itself that may shape its future, such as policy, economic, and social factors. Forecasters who received briefings on these topics provided wider forecast intervals than those receiving no assistance.**

expert elicitation | overconfidence | debiasing | technology forecasting

Forecasts of the future cost and performance of technologies are used in planning a sustainable energy future, as well as to support decision-making in many other public and private contexts (1–6). These forecasts are intended to help decision-makers anticipate future events, avoid surprises, and allocate resources effectively. To that end, technology forecasts should be both as accurate as possible and properly qualified, so that decision-makers know how heavily to rely on them. An ideal forecast provides the correct point value with little associated uncertainty. A realistic forecast provides an interval with a well-calibrated probability of containing the forecasted quantity (7).

However, when forecasts are examined in retrospect, they are often found to be miscalibrated, with the actual value too often falling outside the forecast interval, indicating overconfidence (8–10). Fig. 1 shows probabilistic forecasts of the levelized cost of electricity of photovoltaic technology in 2030, elicited from energy experts in 2009 to 2010 (11, 12). The cost in 2019 (12) was already lower than these experts' lowest assessed values. In a similar study of the efficiencies and costs of emerging photovoltaic technologies, Curtright et al. (10) also found results in which the assessed ranges failed to include future outcomes.

How could these experts have so completely missed the mark? Following standard practice in expert elicitations (13, 14), these experts were warned about research findings that judgments, including forecasts (8, 9), are often seriously overconfident (15–19). Although such warnings are common in expert elicitations, there is little evidence that warnings make a difference (20).

The poor performance of forecasts about photovoltaic technology is not a unique example. Examples of poor retrospective performance of past forecasts, either for future technology cost or performance or for more general topics, such as the predicted cost of infrastructure (21, 22) or energy demand (23, 24), are common. Fye et al. (2) reviewed more than 300 forecasts of the cost, quantity, and performance of a wide variety of technologies (such as robotics, materials, energy, and medical), drawn from scientific papers, industry organizations, and government reports. They defined a successful forecast as one whose event occurred within  $\pm 30\%$  of the forecasted time horizon. They found 38 and 39% success rates for short- and medium-term forecasts (1 to 5 y and 6 to 10 y), respectively, and found only a 14% success rate for long-term (11 to 30 y) forecasts. Albright (25) assessed the accuracy

<sup>a</sup>Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213

Author contributions: T.S., A.D., B.F., and M.G.M. designed research; T.S. analyzed data; and T.S., A.D., B.F., and M.G.M. wrote the paper.

The authors declare no competing interest.

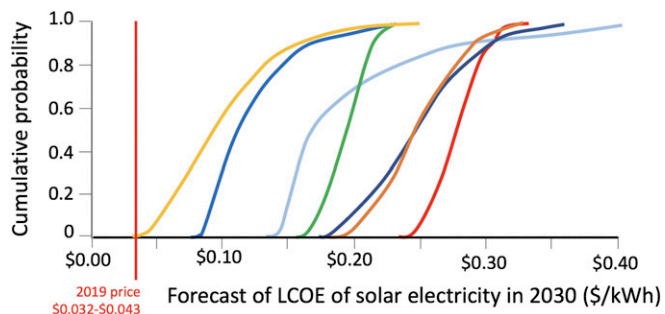
This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

<sup>1</sup>To whom correspondence may be addressed. Email: [granger.morgan@andrew.cmu.edu](mailto:granger.morgan@andrew.cmu.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021558118/-/DCSupplemental>.

Published May 14, 2021.



**Fig. 1.** Elicited cumulative distribution functions obtained from seven energy experts in 2009 to 2010 from a study at Harvard of the likely future costs of photovoltaic technology in 2030 as compared with actual cost in 2019 (11, 12). The vertical red line is the price in 2019 reported by Lazard (12). A study performed by investigators in 2008 at Carnegie Mellon (10) reported expert judgments that were similarly pessimistic about how the future cost of photovoltaic cells would evolve. LCOE, levelized cost of electricity. Adapted from Baker et al. (11).

of Kahn and Wiener's list of "One Hundred Technical Innovations Very Likely in the Last Third of the Twentieth Century" (published in 1967) and found that fewer than 50% were correct (having occurred in the twentieth century). Kott and Perconti (1) did a retrospective of military technology forecasts in which experts assessed the accuracy of statements about a variety of possible future ground warfare technologies over the long term (20 to 30 y), made in 1990 to 2000 for 2020. The forecasts were taken from publications that "ranged broadly from works of fiction to popular articles to organizational internal technical reports and briefs" (1). Kott and Perconti (1) found the accuracy of these forecasts to be 76%.

Fig. 2 displays a number of forecasts for energy consumption. While only a few include interval estimates, none accurately predicted the actual outcomes. Rarely during expert elicitations are experts shown specific examples like these when they are asked to make forecasts. It is also rare that they are given much assistance in thinking systematically about the many policy, economic, and social developments that could shape the future price and performance of a technology. For example, in the elicitations on photovoltaic technology reported in Fig. 1, the experts may not

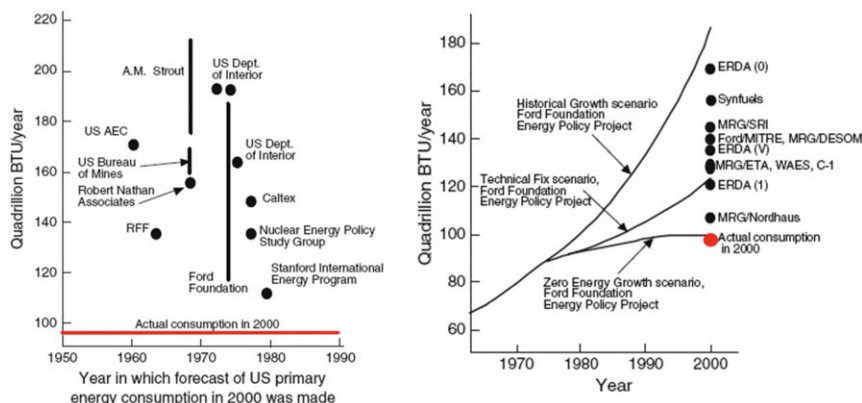
have asked themselves questions about how costs would be affected if many US states implemented renewable portfolio standards or if China decided to make a major commitment to the manufacture of photovoltaic cells. Indeed, in a similar elicitation on photovoltaic technology that we ran at Carnegie Mellon (10), in which the experts were bench scientists working in device-level photocell research, there is no reason to think those experts would have even been in a position to make such assessments without some assistance.

Of course, by offering very wide intervals for their forecasts, experts could be confident in always including the true value. However, that would mean miscalibration in the opposite direction (underconfidence). Experience in both laboratory experiments and practical elicitation tasks has found that people rarely expand their interval estimates widely enough, whatever the warnings provided to them (13, 14, 26, 27).

Some studies have found reduced overconfidence with tasks that encourage people to consider alternative perspectives. Koriat et al. (28) found less overconfidence when respondents listed one reason why an answer might be wrong. Herzog and Hertwig (29) found reduced overconfidence when respondents were asked to produce a second estimate for a question. The Good Judgment Project improved performance by encouraging participants to reflect on the uncertainties and complexities of the processes that they were forecasting (30, 31). Candelise et al. (32) speculated that technical experts could improve forecasts based on historical learning curves by incorporating judgments from experts in related nontechnical fields, such as markets, finance, and business strategy.

Such findings have led us to propose a hybrid approach to probabilistic expert elicitations that would iteratively combine the judgments of technical domain experts and experts knowledgeable about broader issues such as evolving policy environments and technology adoption and diffusion. In such a procedure, all respondents would receive detailed briefings with examples of how poorly many past forecasts have performed.

Here, we report promising preliminary results from a brief intervention that provided examples of poor past forecasts and then asked respondents to consider relevant external developments when assessing a technology's future.



**Fig. 2.** Forecasts (Left) of US primary energy consumption in the year 2000 compiled by Smil (40) as a function of the date on which they were made. Forecasts (Right) of US primary energy consumption for the year 2000 compiled by Greenberger et al. (41) in the early 1980s together with three scenarios developed by the Ford Foundation Energy Project (42). RFF, Resources for the Future; BTU, British thermal unit; AEC, US Atomic Energy Commission; A.M. Strout, Alan M. Strout; ERDA, US Energy Research and Development Agency; MRG, MRG Energy; SRI, SRI International; MITRE, MITRE Corporation; DESOM, Dynamic Energy System Optimization Model; ETA, Energy Technology Assessment; WAES, Workshop on Alternative Energy Strategies; C-1, Case C-1.

## A Pilot Test of the Proposed Strategy

We predicted that technology forecasts would be less overconfident if experts were 1) briefed with specific examples of other experts' poor forecasting performance and 2) prompted to think about external factors that might affect a technology's development. To test the intervention, we asked survey respondents to assess interval estimates for present and future values for aspects of emerging automotive technologies.

In past work, we elicited judgments on the development of advanced nuclear reactors (33) and advanced photovoltaic materials (10). We have a study planned for eliciting forecasts of the future cost and performance of insulated gate bipolar transistors for both silicon and wider band-gap semiconductors. True experts in such cases are so rare, and their time so scarce, that one cannot realistically conduct studies to compare elicitation procedures with them.

The present pilot study adopts a compromise strategy: comparing elicitation procedures in a domain where a relatively large number of individuals have mastered enough technical detail to be considered experts or near experts. We chose another aspect of decarbonization as meeting this criterion: the likely near-term evolution of electric vehicle (EV) technology and autonomous vehicle (AV) technology. Our sample included researchers, engineers, and technically informed enthusiasts.

This study was approved by the Carnegie Mellon University Institutional Review Board. After being shown the informed consent information for the study, participants had to affirm that they were at least 18 y of age, had read and understood the informed consent information, wanted to participate in the study, lived in and were located in the United States, and were at least somewhat knowledgeable about US motor vehicle technology. We recruited 133 respondents from university research centers (such as Carnegie Mellon's Traffic21 laboratory), Reddit, Twitter, the National Academies' Transportation Research Board, the Future of Transportation e-newsletter, and advertisements on [AutomotiveNews.com](http://AutomotiveNews.com) and in SAE International's eSource newsletter. After completing the survey, participants could enter a raffle to win 1 of 10 \$50 gift cards. Data were collected in November and December 2019.

Respondents' task was to assess 80% interval estimates for the current value and the forecasted value in 2 y for 10 questions related to EVs and AVs. If the intervention succeeded, those intervals would be wider and more often include the actual value, once it is known, indicating reduced overconfidence.

To create a broadly applicable intervention, we developed two recorded video briefings. All respondents saw a short (1-min) initial video that introduced the study and explained and illustrated 80% interval estimates. They were also given two practice questions (without receiving answers) to familiarize themselves with the task and the user interface. Respondents were then randomly assigned by the Qualtrics randomizer function to one of three groups: control, one-briefing, or two-briefing. The one-briefing treatment group saw a 2-min video briefing on overconfidence bias. The two-briefing treatment group saw the same briefing, followed by a second 6-min briefing on the poor

performance of past forecasts and advice on how to consider policy, economic, and social factors that might be relevant to their forecasts. The control group saw only the 1-min video explaining the task before providing their interval estimates.

The first briefing (shown to both the one-briefing group and the two-briefing group) described overconfidence, with two illustrative examples. The first was from Soll and Klayman (17) and showed that only 48% of the 80% interval estimates contained the answer to general knowledge trivia questions. The second example was a histogram of surprise indices that summarized results from many studies that found overconfidence in interval estimates (34). Participants were advised to avoid overconfidence by widening their intervals.

The second briefing showed examples of experts' poor performance in past technology forecasts and provided advice on how to think more systematically about policy, economic, and social developments that might affect the forecasts. The examples of poor past forecasts included predictions of the price of solar electricity (11, 12), the pace of technology development (2), and investment returns (8). Participants were shown examples of policy, economic, and social factors that might affect the future of EVs and AVs, such as changes in government regulations, gasoline prices, the availability of EV charging stations, AV crashes, and ride-sharing. They were also encouraged to think of other factors they could add to this list. The video described the implications of increased and decreased gas prices for EV sales. Participants were, once again, advised to avoid overconfidence by widening their intervals.

After viewing their briefings, respondents provided interval estimates for 10 statistics regarding the current state of EV and AV technologies and then provided interval estimates for the same 10 statistics 2 y in the future. For example, they were asked the following questions: "Currently, what is the longest Environmental Protection Agency-rated range for a battery electric vehicle?" "At the end of 2021, what will be the longest Environmental Protection Agency-rated range for a battery electric vehicle?" Of the 10 questions, 6 were about EVs and 4 were about AVs. Questions were asked about the average cost of an EV battery pack, the number of companies with permits to test AVs on public roads in California, the number of public fast-charging outlets in the United States, and other topics. The questions were designed to have publicly available answers, in order to assess which intervals contained the answers and which did not. All respondents saw the questions in the same order.

[SI Appendix](#) provides details about the survey design, including the instructions, briefings, and question wording.

## Results from the Pilot Test

We hypothesized that both briefings would reduce overconfidence in the forecasts, with greater reductions in the treatment group that saw two briefings. These differences would be seen in the widths of the interval estimates and the proportion of intervals that contained the actual value.

Following our preregistered analysis plan (available at <https://osf.io/vxdmb/>), we conducted pairwise two-sample permutation tests for each question in order to determine whether there were

**Table 1. Differences between the three treatment groups when performing the post hoc analysis using a multilevel model on the forecast interval widths**

	Control vs. one-briefing	Control vs. two-briefing	One-briefing vs. two-briefing
Including outliers	$P = 0.002$ (df = 88, $t = 3.16$ )	$P < 0.001$ (df = 86, $t = 5.16$ )	$P = 0.045$ (df = 86, $t = 2.03$ )
Excluding outliers	$P < 0.001$ (df = 88, $t = 3.41$ )	$P < 0.001$ (df = 86, $t = 5.57$ )	$P = 0.030$ (df = 86, $t = 2.20$ )

df, degrees of freedom.

significant differences between the groups (*SI Appendix* has details). Compared with the control group, the one-briefing group had significantly wider intervals for 4 of the 10 forecasting questions (using  $\alpha = 0.05$ ) and 1 of the 10 forecasting questions after Bonferroni adjustment (using  $\alpha = 0.0008$ ). Compared with the control group, the two-briefing group had significantly wider intervals for 8 of the 10 forecasting questions (using  $\alpha = 0.05$ ) and 4 of the 10 forecasting questions after Bonferroni adjustment (using  $\alpha = 0.0008$ ). There were no significant differences in interval widths between the one-briefing and two-briefing groups using the  $\alpha = 0.05$  level. Thus, respondents in both treatment groups sometimes provided wider intervals than respondents in the control group.

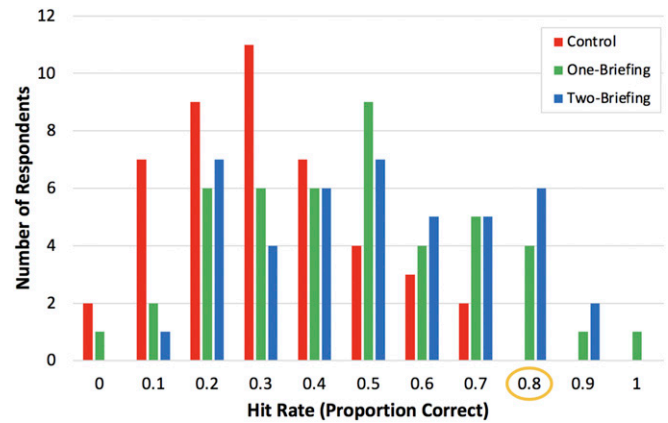
To increase statistical power, we conducted a post hoc multi-level model with per-person random effects to account for the repeated questions answered by each participant (*SI Appendix* has details). Because this model assumes normal errors, we applied a  $\log_{10}(x + 1)$  transform to the forecast interval widths and then standardized the data for each question separately (across the three groups) by subtracting the mean and dividing by the SD of the transformed intervals.

As seen in Table 1, the forecast intervals were significantly wider for both briefing groups than for the control group. They were also significantly wider for the two-briefing group than for the one-briefing group. That pattern appeared when analyzed both with and without outliers (observations more than three SDs from the log-standardized mean).

Fig. 3 shows the distributions of hit rates for judgments of current values, defined as the proportion of each respondent's intervals that contained the correct answer. For appropriately confident respondents, 8 of 10 answers should fall inside their 80% interval estimates. One-sample *t* tests showed that the hit rates for all three groups were significantly lower than 0.8, indicating overconfidence (all  $P < 0.001$ ). However, both treatment groups had significantly higher hit rates than the control group (one-way ANOVA and Tukey's Honestly Significant Difference test;  $P = 0.001$  for one-briefing;  $P < 0.001$  for two-briefing); they were not significantly different from one another. Thus, the treatment groups were less overconfident than the control group, but were still overconfident. At the end of 2021, we will examine the hit rates for the 10 forecasts.

Instructions can induce demand effects (35), encouraging participants to provide what they perceive as expected responses. Here, we sought instructions that balanced the conflicting demands to provide interval estimates that were not too narrow, hence overconfident, and not too wide, hence uninformative. In formulating those instructions, we were guided by previous studies (27, 36, 37) finding that warnings about overconfidence (e.g., "expand those extreme fractiles") have had little effect, whereas asking people why favored answers might be wrong (28, 38) or to "actively entertain more possibilities" (31) can have some effect. Indeed, these findings motivated our asking respondents to list reasons in our second briefing. In addition, when extremely wide intervals were excluded as outliers from the data analysis, the results remained the same (the second row of Table 1).

We find the results of this pilot encouraging. They suggest that a brief, practical intervention, using two video briefings, could



**Fig. 3.** Histogram showing the distributions of hit rates for the current value questions in each group. The appropriately confident hit rate is 0.8 (circled in yellow). Respondents with hit rates below 0.8 are overconfident.

reduce overconfidence in forecasts of technology development. As with any complex intervention, we cannot isolate the effects of specific elements. Based on the behavioral research guiding the briefings, we believe that they managed to guide experts gently to reflect more fully on what they probably already knew about policy, economic, and social factors.

### A Path Forward

When there is a need to forecast the future evolution of an early-stage technology and especially when historical data are scarce, good expert elicitations can inform decision-making. There is evidence in the literature that overconfidence might be reduced and the accuracy of forecasts might be improved if the elicitation process integrates a variety of broader contextual information. However, the literature also makes clear that past efforts to overcome ubiquitous overconfidence have met with limited success. Hence, one should be cautious in judging just how much improvement may be possible.

The results reported here, achieved with simple one-time video-briefing interventions, suggest that stronger effects could be achieved with a hybrid elicitation process that involves iterative interaction between technology experts and experts who are knowledgeable about external policy, economic, and social factors. Facilitating such interaction could help experts produce more accurate and less overconfident assessments of technology futures. We plan to conduct such an assessment and hope that others will do the same.

**Data Availability.** Data, code, and other materials are available on Open Science Framework (<https://osf.io/e2b8c/>) (39).

### Acknowledgments

This work was supported by Grant G-2020-12616 from the Alfred P. Sloan Foundation, by the Center for Climate and Energy Decision Making through Cooperative Agreement SES-0949710 between the NSF and Carnegie Mellon University, and by the Hamerschlag Chair and other academic funds from Carnegie Mellon University.

- 1 A. Kott, P. Perconti, Long-term forecasts of military technologies for a 20–30 year horizon: An empirical assessment of accuracy. *Technol. Forecast. Soc. Change* **137**, 272–279 (2018).
- 2 S. R. Fye, S. M. Charbonneau, J. W. Hay, C. A. Mullins, An examination of factors affecting accuracy in technology forecasts. *Technol. Forecast. Soc. Change* **80**, 1222–1231 (2013).

- 3 E. J. Sylak-Glassman, S. R. Williams, N. Gupta, *Current and Potential Use of Technology Forecasting Tools in the Federal Government* (Institute for Defense Analyses, Science and Technology Policy Institute, 2016).
- 4 Committee on Defense Intelligence Agency Technology Forecasts and Reviews, National Research Council, *Avoiding Surprise in an Era of Global Technology Advances* (The National Academies Press, 2005).
- 5 L. H. Kaack, J. Apt, M. G. Morgan, P. McSharry, Empirical prediction intervals improve energy forecasting. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8752–8757 (2017).
- 6 Energy Information Agency, Annual Energy Outlook Retrospective Review: Evaluation of AEO2020 and Previous Reference Case Projections (2020). <https://www.eia.gov/outlooks/aeo/retrospective/>. Accessed 11 January 2021.
- 7 B. Fischhoff, P. Slovic, S. Lichtenstein, Knowing with certainty: The appropriateness of extreme confidence. *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 552–564 (1977).
- 8 I. Ben-David, J. R. Graham, C. R. Harvey, Managerial miscalibration. *Q. J. Econ.* **128**, 1547–1584 (2013).
- 9 K. Jain, K. Mukherjee, J. N. Bearden, A. Gaba, Unpacking the future: A nudge toward wider subjective confidence intervals. *Manage. Sci.* **59**, 1970–1987 (2013).
- 10 A. E. Curtright, M. G. Morgan, D. W. Keith, Expert assessments of future photovoltaic technologies. *Environ. Sci. Technol.* **42**, 9031–9038 (2008).
- 11 E. Baker, V. Bosetti, L. D. Anadon, M. Henrion, L. Aleluia Reis, Future costs of key low-carbon energy technologies: Harmonization and aggregation of energy technology expert elicitation data. *Energy Policy* **80**, 219–232 (2015).
- 12 Lazard, Lazard’s Levelized Cost of Energy Analysis, Version 13.0 (2019). <https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2019/>. Accessed 31 January 2020.
- 13 M. G. Morgan, Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7176–7184 (2014).
- 14 R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science* (Oxford University Press, 1991).
- 15 S. Lichtenstein, B. Fischhoff, L. D. Phillips, “Calibration of probabilities: State of the art to 1980” in *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, A. Tversky, Eds. (Cambridge University Press, 1982), pp. 306–334.
- 16 J. B. Soll, K. L. Milkman, J. W. Payne, “A user’s guide to debiasing” in *The Wiley Blackwell Handbook of Judgment and Decision Making*, G. Keren, G. Wu, Eds. (John Wiley & Sons, 2015), vol. 2, pp. 924–951.
- 17 J. B. Soll, J. Klayman, Overconfidence in interval estimates. *J. Exp. Psychol. Learn. Mem. Cogn.* **30**, 299–314 (2004).
- 18 H. R. Arkes, C. Christensen, C. Lai, C. Blumer, Two methods of reducing overconfidence. *Organ. Behav. Hum. Decis. Process.* **39**, 133–144 (1987).
- 19 R. A. Block, D. R. Harper, Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organ. Behav. Hum. Decis. Process.* **49**, 188–207 (1991).
- 20 K. L. Milkman, D. Chugh, M. H. Bazerman, How can decision making be improved? *Perspect. Psychol. Sci.* **4**, 379–383 (2009).
- 21 B. K. Sovacool, A. Gilbert, D. Nugent, An international comparative assessment of construction cost overruns for electricity infrastructure. *Energy Res. Soc. Sci.* **3**, 152–160 (2014).
- 22 B. Flyvbjerg, M. S. Holm, S. Buhl, Underestimating costs in public works projects: Error or lie? *J. Am. Plann. Assoc.* **68**, 279–295 (2002).
- 23 Energy Information Administration, *Annual Energy Outlook (AEO) retrospective review: Evaluation of AEO 2018 and previous reference case projections* (US Energy Information Administration, Washington, DC, 2018).
- 24 B. C. O’Neill, Accuracy of past projections of US energy consumption. *Energy Policy* **33**, 979–993 (2005).
- 25 R. E. Albright, What can past technology forecasts tell us about the future? *Technol. Forecast. Soc. Change* **69**, 443–464 (2002).
- 26 M. G. Morgan, M. Henrion, M. Small, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, 1990).
- 27 M. Alpert, H. Raiffa, “A progress report on the training of probability assessors” in *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, A. Tversky, Eds. (Cambridge University Press, 1982), pp. 294–305.
- 28 A. Koriat, S. Lichtenstein, B. Fischhoff, Reasons for overconfidence. *J. Exp. Psychol. Hum. Learn.* **6**, 107–118 (1980).
- 29 S. M. Herzog, R. Hertwig, Think twice and then: Combining or choosing in dialectical bootstrapping? *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 218–232 (2014).
- 30 D. A. Moore et al., Confidence calibration in a multiyear geopolitical forecasting competition. *Manage. Sci.* **63**, 3552–3565 (2017).
- 31 B. Mellers et al., Psychological strategies for winning a geopolitical forecasting tournament. *Psychol. Sci.* **25**, 1106–1115 (2014).
- 32 C. Candelise, M. Winkler, R. J. K. Gross, The dynamics of solar PV costs and prices as a challenge for technology forecasting. *Renew. Sustain. Energy Rev.* **26**, 96–107 (2013).
- 33 A. Abdulla, I. L. Azevedo, M. G. Morgan, Expert assessments of the cost of light water small modular reactors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9686–9691 (2013).
- 34 M. G. Morgan, *Theory and Practice in Policy Analysis: Including Applications in Science and Technology* (Cambridge University Press, 2017).
- 35 M. T. Orne, On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am. Psychol.* **17**, 776–783 (1962).
- 36 J. K. Adams, P. A. Adams, Realism of confidence judgments. *Psychol. Rev.* **68**, 33–45 (1961).
- 37 G. F. Pitz, Subjective probability distributions for imperfectly known quantities. *Knowledge and Cognition*, 29–41 (1974).
- 38 P. Slovic, B. Fischhoff, On the psychology of experimental surprises. *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 544–551 (1997).
- 39 T. Savage, Motor vehicle technology forecasting survey. Open Science Framework. <https://osf.io/e2b8c/>. Deposited 30 April 2021.
- 40 V. Smil, Perils of long-range energy forecasting: Reflections on looking far ahead. *Technol. Forecast. Soc. Change* **65**, 251–264 (2000).
- 41 M. Greenberger, G. D. Brewer, W. W. Hogan, M. Russell, *Caught Unawares: The Energy Decade in Retrospect* (Ballinger, 1983).
- 42 The Ford Foundation, *Energy Policy Project, “Exploring Energy Choices: A Preliminary Report”* (The Ford Foundation, 1974).