TITLE:

Quantifying the effects of expert selection and elicitation design on experts' confidence in their judgments about future energy technologies

(this part added)

Gregory F. Nemet,[a] Laura Diaz Anadon,[b] Elena Verdolini[c]

a. La Follette School of Public Affairs, University of Wisconsin–Madison WI, USA; Nelson Institute Center for Sustainability and the Global Environment (SAGE), University of Wisconsin–Madison WI, USA

b. John F. Kennedy School of Government, Harvard University, MA, USA; Department in Science, Technology, Engineering and Public Policy, University College London, UK

c. Centro Euro-Mediterraneo sui Cambiamenti Climatici and Fondazione Eni Enrico Mattei, Italy

In the acknowledgements to include: "All authors contributed equally to this article"

ABSTRACT:

Expert elicitations are frequently used to characterize uncertain future technology outcomes. However, their usefulness is limited, in part because: estimates across studies are not easily comparable; choices in survey design and expert selection may bias results; and over-confidence is a persistent problem. We provide quantitative evidence of how these choices affect experts' estimates. We harmonize data from 16 elicitations, involving 169 experts, on the 2030 costs of 5 energy technologies: nuclear, biofuels, bioelectricity, solar, and carbon capture. We estimate determinants of experts' confidence using expert characteristics, survey design, and public R&D investment levels on which the elicited values are conditional. We find that when experts respond to elicitations in person (vs. online or mail) they ascribe lower confidence (larger uncertainty) to their estimates, but more optimistic assessments of best-case (10^{th} percentile) outcomes. The impact of expert affiliation (government, private sector, or academic) and geography (US or EU) varies by technology. Academics and public sector experts generally express lower confidence than private sector experts. EU experts are more confident than US experts, driven mainly by biofuel costs. Higher R&D spending can increase uncertainty rather than resolve it, but it consistently reduces best-case cost estimates. These results indicate the source, direction, and size of bias in a large sample of energy technology elicitations. They also point to the technology specificity of some of the effects. We discuss ways in which these biases should be seriously considered in interpreting the results of existing elicitations and in designing new ones.

Key Words: expert elicitations, uncertainty, energy technologies, heuristic biases, survey design

MAIN TEXT:

1 INTRODUCTION

Policy makers addressing science and innovation issues frequently confront the challenge of making decisions that affect the development of technologies. Their decisions rely on explicit[(1)] or implicit[(2)] characterizations of the anticipated cost and performance of specific technologies. However, both technology outcomes and consequent social impacts are notoriously difficult to predict, because technologies change over time, often in ways that diverge from historical trends[(3)]. Expert elicitations allow analysts to make use of information collected from experienced professionals about the future of specific technologies that may not be available from other data sources. Importantly, expert elicitations also provide measures of uncertainty associated with the central estimates. Protocols for data collection are designed to reduce biases and encourage considered judgments[(4-6)]. These data can provide crucial information for policy design. Indeed, although in the past they were used mainly in the private sector[(7)] expert elicitations are increasingly used in policy making, starting in the 1970s with the U.S. Environmental Protection Agency (EPA), and now with at least five other federal agencies and international organizations using them[(8, 9)].

Because of the substantial externalities and the very long time horizons inherent in energy systems, policy makers are particularly interested in using elicitations to inform decisions around public funding to support innovation in energy technologies. In the U.S, for instance, the National Research Council published a strong recommendation that the U.S. Department of Energy (DOE) use expert elicitations to inform for their R&D allocation decisions to probabilistically characterize the expected outcomes of R&D investments[(9)]. Over the past decade, optimism about the potential for expert elicitations to inform public decisions stimulated the launch of more than twenty expert elicitation studies on several important energy technologies. Some ask questions about future cost and performance conditional on public R&D investments and a subset of these elicitations are listed in Table I. However, differences in protocol design (metrics, assumptions, timeframes, methods for administering the surveys) and in the backgrounds of experts selected (institutional affiliation and nationality) make it hard for the analyst or policy maker to compare the results from these studies on a consistent basis. This issue of how to utilize and learn from the existing expert elicitations for future elicitations extends beyond the debate that recently took place in this journal regarding how, or whether, to derive consensus from them[(10-14)]. A key question is whether differences between studies themselves affect the elicitation results, both in terms of the distribution and central estimates.

Morgan provided a recent review of how to think about selecting experts, in part to reduce bias[(7)]. However, despite a robust discourse about how to make use of elicited data, the literature does not yet include a quantitative assessment of how differences in study design and expert selection affect elicited values[(15-21)]. Two recent articles provide first steps in this direction by focusing on elicitations in nuclear and solar technologies [(22, 23)]. In this paper, we expand the analysis by assessing the roles of expert and survey characteristics on the experts' uncertainty range and best-case estimates in five key energy technologies – nuclear fission, biofuels, bioelectricity, solar, and carbon capture in coal power plans. We address the research question: *how do elicitation design and expert selection choices affect the confidence of experts' responses to elicitations of future technology outcomes?*

Our results provide a clearer sense of how much uncertainty exists in anticipated technology outcomes—and how elicitation design can affect that uncertainty—and in doing so they help interpret the results of existing elicitations and provide evidence to improve the design of future elicitations. The next section provides an overview of the data we use and our approach to empirically addressing the research question above. Section 3 presents the results. Section 4 includes a discussion of implications for policy makers, particularly on debiasing, geographic considerations, and the effect of R&D on uncertainty. This discussion mentions ways in which the results serve as a basis for improving policy making— for example, in government agencies such as the U.S. Dept. of Energy, multiple Congressional committees, and the European Commission—by facilitating the interpretation of existing energy technology elicitations.

2 APPROACH

2.1 The elicitation data

We collect and standardize data from 16 expert elicitations of expected technology costs in 2030.[1] These studies include five energy technologies: nuclear fission power, biofuels, bioelectricity, solar photovoltaic power, and carbon capture and storage (CCS) for power generation.[2] The future costs of these technologies are important for policymakers facing decisions on climate change as well as on energy supply[(24, 25)]. As shown in Table I, the diverse characteristics of these 16 elicitations conducted over a relatively short period of time provide an unusual opportunity to study whether selection and survey design affect experts' estimates. Similar to previous work on infectious disease and marine ecology, our general approach is to regress elicitation responses on study and expert characteristics[(21)].

The expert elicitations in Table I

Table I include each expert's estimates of future costs at three points on their tacit probability distribution of future costs. Nine of the 178 experts were included in two studies as opposed to just one, and thus the

---

[1] Throughout, we refer to `costs' as the cost in 2030 to an adopter of the technology producing useful energy.
[2] Three additional studies were ultimately dropped from the analyses due to lack of specification of public R&D funding conditions.

dataset includes 169 unique experts. In addition to having a very small number of overlapping experts, the overlaps happened primarily in the technology areas with the largest number of experts (nuclear and solar), with only two experts overlapping on two other surveys. Thus, treating repeat experts as individual responses in two different surveys is not expected to bias the results.

Each study made substantial efforts to reduce bias and overconfidence. For example, in the online Harvard nuclear study[(26)], after reviewing background information on nuclear costs and public R&D budgets, experts engaged in modules on reducing bias, overconfidence, and estimating percentiles. The module on overconfidence includes a discussion and a widely used figure illustrating the poor ability of experts to estimate their confidence around the speed of light[(27)]. To reduce overconfidence, experts were asked to provide the 90th and 10th percentiles before the 50th percentile and were instructed to review their answers, imagining alternate scenarios wherein the true value is outside the range they have provided through six steps. The module on bias cautions experts to be aware of the tendency of experts to be biased due to familiar experiences and subjects. The percentiles section included an explanation of how to think about percentiles, giving examples of what the 10th, 90th, and 50th percentiles represent, as well as a Figure that mimicked the interactive graph experts would generate in their answers later on. The 10th percentile (P10) is the expert's estimate of a "best case" outcome, the 50th percentile (P50) is the "most likely" outcome, and the 90th percentile (P90) is the other extreme, the worst case they can imagine.[3]

[3] One exception were the UMass studies, which asked experts about probabilities of particular goals being met and then converted the answers to percentiles, as mentioned below and detailed in the SI.

We obtained these elicitation data from the original authors and converted them into common units of levelized energy cost (LEC) in units of *2010$/kWh*.[4] Even though they all produce energy, the technologies are sufficiently heterogeneous that we use a slightly different system boundary in calculating each LEC: solar includes the full levelized costs of each unit of electricity; bioelectricity uses the non-fuel levelized cost of electricity; biofuel uses the non-fuel levelized costs of each unit of energy; nuclear uses the levelized capital costs of electricity; and CCS uses the levelized costs of the additional capital costs for each unit of electricity produced. In the cases of solar data from Harvard and CMU, and the CCS data from UMass, a simple techno-economic model was used to produce LEC estimates from elicited components of the LEC.

Each study asked questions about future costs under one or more levels of public R&D funding. Table I shows how the R&D scenario terminology used in the various studies was translated into three the common R&D scenarios we use in this paper: Low, Mid, and High. The Low R&D scenario is largely consistent with a business as usual public R&D funding case for that particular technology in the particular region in which the elicitation was conducted. The Mid R&D scenario is consistent with a significant increase in R&D funding across surveys. In the case of Harvard this significant increase was the expert's "recommended" R&D levels, in the FEEM surveys (with the exception of the nuclear FEEM survey, which was consistent with Harvard's definition) the Mid R&D scenario was a "50% increase in the BAU funding." The High R&D scenario was consistent with a very large increase in R&D funding. In the case of the CMU data this was consistent with 10 times the "BAU" R&D level. The High R&D scenario in the Harvard survey was 10 times the "recommended" R&D scenario. In the FEEM surveys, the High R&D scenario

---

[4] Levelized costs are common in energy and involve summing amortized up-front costs and variable costs and then dividing by units of energy produced.

was 2 times the BAU level. Finally, the UMass surveys provided funding amounts and used the terms "Low", "Mid" and "High" R&D scenarios directly, and we use the same terminology here[5].

Figure 1 and Figure 2 show the elicited point estimates (at the 10th, 50th, and 90th percentiles) for the five energy technologies under the each of the three R&D scenarios separately for each of the studies. Table II shows descriptive statistics for the individual participant data of the 16 expert elicitations. For the binary variables, the average indicates the fraction of the observations representing various technologies, types of experts, R&D levels, etc. Explanations of technology sub-type, study characteristics, and R&D scenarios on which the estimates are conditional are documented in the SI.

2.2 Dependent variables

In this study, we focus on two dependent variables. First, given the growing interest in consideration of uncertainty in science policy decisions[9], and the vast literature on the cognitive biases in the subjective assessment of probabilities[28, 29], we focus here on experts' confidence around central estimates. The uncertainty range ("*Urange*", henceforth) is defined as the difference between the 90th and 10th percentile divided by the 50th:

$$\text{Urange} = (\text{P90}-\text{P10})/\text{P50} \qquad (1)$$

It measures the percentage variation from each expert's median estimate within each of the R&D scenarios. Figure 3 shows probability density functions of the Uranges for all data in our sample, both overall and for each technology. Second, because the left tail of the cost distribution (low costs) is of particular interest to

[5] The solar UMass survey is an exception, since it only has two R&D scenarios.

policy makers—and because it adds insight on what is driving changes in Urange—we also include models in which we use the best case estimates (P10) as the dependent variable.[6]

Note that since Urange is a normalized metric, it can be pooled for all technologies, even if the standardized costs measure different parts of the technology. Hence, for this metric we present both pooled results and technology specific to show the robustness of our results to different assumptions. Conversely, elicited percentile metrics can be meaningfully compared only within technology due to the differences in what is included in the standardized costs.

2.3 Independent variables

We assess the extent to which the following four aspects may systematically affect the uncertainty range and the best case estimates: (a) technology characteristics, (b) expert elicitation survey design, (c) expert characteristics, and (d) R&D investment levels on which the elicited values are conditional. We selected these factors because the available qualitative literature on the subject suggests they have an impact on elicited values.

Uncertainty ranges and best-case estimates across studies can vary due to the diversity in the technologies considered. Figure 1, for instance, highlights the wide diversity of expert opinion on the 50th percentile 2030 estimates of solar and nuclear technologies under the low funding scenario, highlighting the large differences within solar and nuclear but also across both technology areas. Possible explanations for such differences across technology areas include the maturity of a given technology, the extent to which learning-

[6] In the SI we also include results on the relationship between the variables of interest and the 50th percentile estimate.

by-doing has improved costs in the past, the number of technological paths that have already been explored, and the specific efficiency of each technological path. In our pooled analysis of Urange, we include dummies for *bioelec, biofuel, nuclear, solar*, and *ccs* to capture differences in average Urange values across technologies, while such differences are already implicit in the technology specific regressions.

The literature on elicitation has looked at the differences in the design of elicitation protocols, highlighting in particular the importance of the expert selection phase and of the method by which the survey is administered (in person, via mail, or internet)[15-20, 30]. It devotes significant attention to issues such as the optimal number experts and the careful sampling for expert selection. It also points to the advantages of in-person elicitations[5, 17]; during in-person interviews the researcher can devote more time to "debiasing" and can ask follow-up questions that prompt experts to consider a wider range of possible outcomes. However, in-person elicitations are far more costly and time-consuming, particularly for investigators. Researchers address this trade-off by carefully designing mail or online elicitation protocols. Specifically, they provide very detailed background information to reduce bias, and often resort to interactive tools such as visualization software to provide timely feedback to the expert and given him/her opportunities to correct their answers. To date, there is no quantitative evidence indicating whether, even in light of careful protocol design, there is a systematic difference in the level of confidence between experts involved in the two types of elicitations. We code each elicitation using the binary variable, *inperson*, which assumes the value of 1 for in-person interviews, and zero otherwise.

Similarly, expert background (e.g., institutional affiliation and country) is likely to affect cost estimates[30, 31]. Moreover, elicited data is likely to be subject to availability and anchoring heuristics associated with experts' environment and experiences[29]. Experts from industry are likely exposed to different information sets than those from academia and the public sector, since they are likely working at different stages of technology development, participating in different conferences, and interacting with different colleagues. Along the same lines, experts living in different geographical areas may have varying experiences with cost overruns, public opposition, and policy support. We code each expert as based in *academia*, the *private*, or the *public* sector. Furthermore, since each expert was asked to provide estimates for the region in which they work, we code this information as *EU* or *US*.

Finally, the suggested public R&D investment levels included in the elicitation can have an impact on uncertainty. The direction of this effect is largely an empirical question. In fact, experts may make estimates with wider uncertainty ranges (lower confidence) under higher R&D investment assumptions if they have difficulties imagining outcomes that are far away from the actual state of the world. However, experts could also have narrower uncertainty ranges, if they expect higher R&D to solve technical issues that are unresolved under scenarios with lower R&D investments. Our question related to the impact of R&D on the uncertainty about the future cost of energy technologies is similar to that in Zickfeld et al. (2010) who asked experts about the role of additional research in reducing uncertainty around the global temperature response to specific trajectories of radiative forcing[(32)]. We include variables for the levels of public R&D (RD) in million 2010 US$/year, as well as binary variables representing bins of low (*RD_lo*), medium (*RD_mid*), and high (*RD_hi*) R&D.

Table I shows the assignment of R&D funding levels in the different studies to these three bins.

2.4 Estimation approach

Given the heterogeneity across studies, we use regressions to control for confounding factors at the individual expert level and for treatment differences between studies[(33, 34)]. Our basic specification is as follows:

$$\ln\left(Y_{itpr}\right) = \alpha + \beta \ln(S) + \gamma \ln(T) + \delta \ln(E) + \theta \ln(R) + \vartheta_{it} + \varepsilon_{iptr} \qquad (2)$$

where *i* indicates expert, *t* technology, *p* a given sub-technology (*subtech*), and *r* the specific R&D scenario (or level) on which the elicited metric is conditional. We set our dependent variable, *Y* as Urange to measure confidence and as P10 to assess best case outcomes. *S* is a dummy variable equal to one if the survey was conducted in person; *T* are dummy variables indicating the technology focus of the specific elicitation, with solar as the reference category; *E* are dummy variables indicating the expert was from academia or the public sector, with private sector being the reference category; and *R* are variables indicating the R&D scenario with which each estimate is associated. Summary statistics are presented in

Table I and Table II. We propose two specifications for *R*. In the first, dummy variables indicate medium and high funding (with business-as-usual funding being the reference). In a second specification, we use the continuous R&D variable (in constant 2010 dollars) to test the robustness of the binned specification and explore the possibility of diminishing marginal returns by including the squared term in the regression.

We use random effects models in which each observation is a combination of expert and sub-technology, observed over different R&D scenarios, to control for expert effects, in addition to the other control variables related to survey design, expert selection, and R&D investment level[23]. Standard errors are clustered at the level of expert and subtechnology.

3 RESULTS

3.1 Estimating Uncertainty

We use equation 2 to estimate uncertainty range, first by pooling data for all technologies (Models 1 and 2), and then for each of the five technologies separately (Table III).

3.1.1 Relationship between Urange and survey design characteristics

Models 1 and 2 in Table III—pooling all technologies and including technology and random effects—indicate that elicitations conducted in person have uncertainty ranges that are 33% greater than those that were conducted online or over the mail, on average and *ceteris paribus*. This positive and statistically significant result (at a 1% level) is robust to conducting technology specific analysis (as shown in Models 3-7), the only exception being the positive but not statistically significant coefficient for bioelectricity (Model 5). These results indicate that when experts respond to elicitations on the future costs of energy technologies in person, they ascribe lower confidence (larger uncertainty) to their estimates than when responding via mail or internet. In the SI we include a figure showing the effect on Urange of shifting to in-person elicitations (Figure S-3).

3.1.2 Relationship between Urange and expert characteristics

In Model 1 of Table III, the coefficient associated with experts in academia suggests that, on average and across technologies, academics' Uranges are roughly 12% greater than those in the private sector. Public sector experts are also associated with higher Uranges than those in the private sector, but the estimate is statistically significant only when using a continuous R&D variable (Model 2). Looking at the technology specific regressions it is clear that both the magnitude and the precision of the estimate are driven by nuclear experts (Model 4). Overall, EU experts appear, on average, more confident, with an uncertainty range that is 12% lower than that of US experts (Model 1 and Model 2). In this case, the magnitude and significance is mostly attributable to experts in biofuel technologies (Model 5). Hence, differences in elicited estimates due to expert background and geographic area are technology specific. This difference in Urange depending on expert characteristics could be explained by availability biases and by differences in local permitting and regulatory costs.

3.1.3 Relationship between Urange and R&D variables

Table III indicates that, overall, the three R&D scenarios upon which the cost estimates are conditional do not have a significant impact on experts' confidence. However, when looking at the technology-specific results, the higher R&D scenarios are associated with more uncertain estimates (lower confidence) around future costs for solar, and less uncertain estimates (higher confidence) for biofuels. These effects may be due to increasing R&D investments pushing researchers to expand the range of technological possibilities for solar, whereas in biofuels experts may be more certain about the possibilities due to a focus on particular technical bottlenecks to overcome. In this respect, note that solar is the technology for which R&D has the largest effect on median (P50) future costs (see Table S-5 in the SI). Conversely, for biofuels the mid and high R&D scenarios are associated with the lowest uncertainty ranges.

3.1.4 Relationship between Urange and technology categories

Pooled Models 1 and 2 in Table III show that, on average, Urange in solar (the omitted technology dummy) is statistically different from those in the other four technology categories. Urange is on average 17%, 19%, 18% and 63% higher for nuclear, bioelectricity, biofuels, and CCS experts, respectively, compared to solar. That different technologies are associated with different perceptions of uncertainty is not surprising, but to the best of our knowledge this is the first empirical assessment of the extent to which experts' confidence is greater in some technologies versus others. In the specific case of the technologies considered here, the small number of new constructions in both nuclear and CCS might be a source of their higher uncertainty.

We acknowledge that the range of observed characteristics we control for in our regression is unlikely to account for all variation beyond the core technical judgments we are attempting to elicit. For example, in the SI we discuss our attempt to evaluate the impact of two additional elicitation variables of interest to the meta-analysis literature: the year in which the study was conducted and whether or not the results were published in the peer-reviewed literature. We were however unable to determine their impact in a robust manner. First, all elicitations were carried out only a few years apart, providing very little variation. However, we do see that most of the elapsed time between the earliest and latest studies is due to the UMass studies; results in the SI shows that when we drop UMass, we get similar results. Second, in a few cases the year of elicitation was different between different studies, resulting in collinearity issues with other variables. The same is true for the "published" variable vis-a-vis the E.U. and in-person variables for some technologies. We drop the "published" dummy variable because researchers decisions to publish some of the elicitations in a non-peer review outlet were unrelated to the design or implementation of the elicitation.

3.2 Estimating best case outcomes

To add insight on what may be driving the Urange results, we also regress the 10th percentile cost estimate (the best-case outcome) on the independent variables separately by technology (Table IV). The heterogeneity in the results for expert selection variables across the different technologies in the case of P10 is higher than in the case of Urange. Academic experts provided more optimistic P10 estimates for nuclear and CCS, and more pessimistic estimates for biofuels than their industry counterparts (see Models 2, 3, and 4, respectively, in Table IV). Public sector experts provided more optimistic P10 estimates for nuclear and bioelectricity, and more pessimistic estimates for biofuels than their industry counterparts. EU experts are more pessimistic about P10 than their US counterparts in bioenergy and biofuels. This suggests that previous experiences may be more conducive to differences in the sign of perceptions of best cases than on the uncertainty range in the technology areas evaluated in this study. Furthermore, in line with the Urange results, the results from the regressions on the best case outcome confirm that the relationship between expert background and the best-case estimates is technology specific. The finding that for different technologies different types of experts are more optimistic could be explained by availability biases, i.e., by the fact that experts working in the same technology may face different types of information and experiences depending on which sector they are in. For instance, nuclear industry experts in the EU and the US have less ability to work on actual deployment projects when compared to solar industry experts since over the past couple of decades solar deployment in the EU and the US has been much more prevalent than nuclear deployment. This may lead solar industry and academic experts to have a similar sense of what future costs may be (no appreciable differences were found in their estimates of the uncertainty and best case costs). In contrast, nuclear academic experts were systematically more optimistic about future costs and less uncertain when compared to nuclear industry experts.

As expected, higher R&D investments are associated with lower P10 values. This result is statistically significant for all technologies, with the exception of CCS, which has only 18 observations. Note also that the sizes of the coefficients are quite different among the technologies in Models 1-5. That the R&D amounts in each bin vary by technology may explain the differences in the sizes of the coefficients. They may also reflect different beliefs about the impact those R&D investments will have on future outcomes. In particular, the coefficient of the impact of the mid and high R&D scenarios on P10 is largest in the case of solar power. These differences may get at fundamental differences in technological opportunity across technologies[35].

Similarly to the results for Urange, the effect of the in-person variable on P10 is mostly consistent across the five technologies. It is negative and significant in the solar, nuclear and biofuel regressions (those with the largest number of observations) and insignificant for bioelectricity and CCS. Overall, these results suggest that in-person elicitations are likely to be associated with more optimistic P10 estimates compared to mail or internet elicitations. Lower P10 values help explain why in-person is associated with greater uncertainty ranges (lower confidence), as discussed above.

4 IMPLICATIONS

Our analysis provides three main results quantifying: (1) how elicitation design affects expert confidence; (2) how expert confidence varies across by expert background, region, and technology; and (3) how R&D investments affect expert confidence. In addition, we show that expert's best guess is also shaped in important ways (sometimes different to those we showed in the case of the uncertainty) by various variables. Our core finding—that elicitation design affects expert confidence—in combination with the other results, has several implications for policy design, as well as for future elicitation research. The realm of public funding of technology includes an inherent aspect of uncertainty. An implication that runs through the various points in the discussion below is that, in the domain of energy technology, policy analysts and decision makers will have to operate under conditions of more rather than less uncertainty, even if they increase public R&D funding. The resulting need for nuanced and cautious interpretation of elicitations results exacerbates what is already a persistent challenge: how to communicate the results of expert elicitations to policy analysts who need to situate these results into a much more complicated world than a well-structured elicitation instrument could ever control for.

4.1 Minimizing expert over-confidence is costly and valuable

A longstanding challenge in expert elicitation has been to find ways to overcome experts' biases to think too narrowly about possible outcomes—even if many decision makers consider results with high confidence to be more useful than those with low confidence[14]. Given what we know about expert over-confidence and the results from this analysis, consumers of elicitation studies in energy technologies should pay close attention to the *in-person* variable results. That in-person elicitations reduce confidence indicates that on-line techniques are still not yet sufficiently close substitutes for in-person interviews. We think this result is particularly robust for two reasons. First, because the result was obtained even after controlling for other possible variables affecting the uncertainty range, such as the background or geographic focus of the expert. Second, the authors of the online and mail elicitations included here were well aware of overconfidence and made substantial efforts to address it; they provided examples to make experts aware of their potential biases, instructions of steps to follow, and interactive tools to help them visualize their responses (and how they related to each other) allowing experts to adjust responses as they went along. However, a well-trained interviewer can perhaps better convey the importance of thinking about extremes, ask relevant follow up questions, and prod an expert to move beyond glib responses and gut feeling. A possible rival interpretation is selection; experts who are amenable to investing their time in an in-person interview may be more likely to have a broader view of uncertainty. In either case, in-person reduces over-confidence in the results. Note also in Table IV that in-person increased the uncertainty range by reducing the best case (P10), rather than increasing the worst case (P90 in the SI). These results suggest that the mechanism by which in-person addresses over-confidence is by prompting consideration of technological possibilities, rather than in identifying potential obstacles.

In-person interviews are more time consuming when considering the sum of the full set of activities involved in conducting an elicitation: researcher preparation, travel, interview, and post-interview processing, as well as experts' participation and travel (in some cases). Is the value of information worth it? Scaling up efforts to perform more elicitations might ultimately be helped by comparing the benefits of lower confidence to the costs on in-person interviews, as well as to the alternative of improving online elicitations. An important opportunity of future work would be to test the results presented here in an experimental setting. Such work could build on recent work in which EU experts and US experts responded to the same online elicitation tool[(26)]. This level of control would allow researchers to identify statistically significant differences in the answers of experts in both regions and could also be used to assess the effect of question format[(21)].

4.2 Public R&D can increase uncertainty

A potentially disappointing result for policymakers confronting the large uncertainty in future energy technology costs is that R&D investment does not appear to necessarily reduce uncertainty. From a social perspective, public R&D investments in energy are thought to provide multiple public goods: 1) they can improve technology outcomes, but 2) they can also generate information that can, for example, help inform future decisions[36]. Our results suggest that experts expect the former but not the latter. High R&D investments are consistently associated with lower P10 cost estimates, but high R&D scenarios do not affect experts' confidence in the outcomes. These results, for these five technologies, provide a much stronger case that R&D will improve technologies by 2030 than it does that more R&D will clarify expectations about which technologies will be most promising between now and then, given that expectations around a best guess are unlikely to converge in the short term. Instead, the results show that instead of waiting for the uncertainty to be resolved (which R&D may not be able to accomplish in the short term) R&D investment decisions need to fully embrace the large uncertainties (and opportunities) awarded by R&D in the long-term in an option value framework. Given the difficulties in communicating uncertainty to policy makers and the public, supporting R&D investment decisions incorporating (perhaps even larger) uncertainties provides real challenges to R&D program evaluation efforts between now and 2030. The future of key energy technologies is likely to get murkier rather than clearer if the public sector increases support for it. That will be particularly challenging in that higher public R&D budgets are where program evaluation is most beneficial and in which close scrutiny is more likely. On a related note, the results of this study could be utilized in modeling exercises that aim to inform policy design related to energy technologies. These include the Energy Modeling Forum (EMF) and the Intergovernmental Panel on Climate Change (IPCC). The study authors have been interacting with the EMF group on a formal basis and expect that this interaction will continue to help transfer the insights from this study into a range of integrated assessment models, building on other work[37, 38].

4.3 A heterogeneous pool of experts is needed for robust insights

As indicated by the effects of expert affiliation and location, expert selection has the potential to influence results. In specific technologies, US experts were more uncertain about future costs than EU experts and academics are generally more uncertain than their industry counterparts. Provided that all are experts, interpreting this result for designing future elicitations is less straightforward than the in-person result. This could again be related to availability biases—experts not engaged in taking technologies into the market may have more uncertainty regarding what may take to achieve commercialization, including uncertainties related to technology performance at scale. European experts may have lower uncertainty in general because they have had more recent experience with biofuels (a technology for which EU is statistically significant), and other technologies (for which the coefficient is also negative but not significant). In sum, these results provide a strong reminder for expert elicitation researchers, and consumers of them, to rely on a heterogeneous set of experts to limit the bias in the elicited results since experts' environment may affect their access to disparate information and thus their confidence about future technology costs. If we want a broad swatch of expertise, it seems unwise to preferentially include Americans and academics to other types of experts in order to minimize within-expert over-confidence. In some cases, selecting a heterogeneous set of experts may not make a difference, as for instance in the case of the solar elicitations in our sample. However, given our inability to predict *ex ante* whether or not different experts will be associated with estimates that are statistically different, the robust approach requires a broad composition of expert backgrounds to be represented.

4.4 Interpreting local perspectives in a global system

Our results overall show significant variation in expert estimates across regions, even after controlling for other possible confounding factors. Further, we know from other studies that substantial within-region variation exists, even for globally traded and apparently homogenous technologies[39-41]. Three main implications follow. First, when dealing with public policy choices regarding energy technologies, there is no "law of one price"[42]. The variation in best estimates and expert confidence across different technologies in the same region can be large, and such differences need to be factored into choices regarding public R&D portfolio investments and other policies affected by expectations future technology costs. Second, the variation within regions is accompanied by significant variation across regions. This means that energy technology markets, which exhibit some differences in LEC costs today[43] are not expected to have globally equivalent costs in the medium term. As a result, the policy choices and modeling efforts in different regions need to be tailored accordingly. Third, this raises questions about the geographic external validity of expert elicitation exercises. Can insights from existent expert elicitation studies—overwhelmingly conducted in developed countries—be applied to places with different technology and policy contexts, such as those of China or other developing countries? The answer is important because those are the regions with the largest expected energy demand in the decades to come. The differences in expectations about future technology cost across regions imply the need to conduct expert elicitations within those contexts to inform policy decisions strongly shaped by expectations about the evolution of technology.

4.5 Conclusion

Despite the dearth of alternative means by which to estimate future technology outcomes, expert elicitations remain vulnerable to criticisms of being unrepresentative, merely subjective, and based on opinions rather than facts. If elicitations are to be considered important evidence to inform decisions, involving potentially billions of dollars of public funds, they need to be credible, which requires an improved understanding of what determines elicitation outcomes. We suggest that an empirically-based understanding of what drives the range of experts' responses can increase the effectiveness of expert elicitations in supporting policy decisions involving science and innovation. Improved credibility will be socially useful even if—and perhaps especially if—the primary influence of elicitation studies is to broaden policy makers' understanding of what is possible in the future.

REFERENCES

1. Nakicenovic N, Riahi K. An Assessment of Technological Change Across Selected Energy Scenarios. Research Report. IIASA, 2002 May. Report No.: RR-02-005.

2. Nordhaus WD. A Question of Balance: Weighing the Options on Global Warming Policies: Yale University Press; 2008.

3. Gallagher KS, Grubler A, Kuhl L, Nemet G, Wilson C. The Energy Technology Innovation System. Annual Review of Environment and Resources. 2012;37(1):137--62.

4. Hogarth RM. Judgement and choice the psychology of decision Chichester, UK; New York: Wiley; 1987.

5. Morgan G, Henrion M. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge: Cambridge University Press; 1990.

6. Cooke RL. Experts in Uncertainty: Opinion and Subjective Probability in Science. New York: Oxford University Press; 1991.

7. Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences. 2014:201319946.

8. EPA. Expert Elicitation Task Force White Paper. Report. Washington DC: {U.S.} Environmental Protection Agency; Prepared for the Science and Technology Policy Council, 2011.

9. {NRC}. Prospective Evaluation of Applied Energy Research and Development at DOE (Phase Two). Washington: The National Academies Press; 2007.

10. Bolger F, Rowe G. The Aggregation of Expert Judgment: Do Good Things Come to Those Who Weight? Risk Analysis. 2014;35(1):5--11.

11. Bolger F, Rowe G. There is Data, and then there is Data: Only Experimental Evidence will Determine the Utility of Differential Weighting of Expert Judgment. Risk Analysis. 2015;35(1):21-6.

12. Cooke RM. The Aggregation of Expert Judgment: Do Good Things Come to Those Who Weight? Risk Analysis. 2015;35(1):12-5.

13. Winkler RL. Equal Versus Differential Weighting in Combining Forecasts. Risk Analysis. 2015;35(1):16--8.

14. Morgan MG. Our Knowledge of the World is Often Not Simple: Policymakers Should Not Duck that Fact, But Should Deal with It. Risk Analysis. 2015;35(1):19--20.

15. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. Oxford: Addison-Wesley; 1968.

16. Keeney RL, Winterfeldt DV. Eliciting probabilities from experts in complex technical problems. Transactions on Engineering Management. 1991;Vol. 38:191-201.

17. Meyer MA, Booker JM. Eliciting and Analysing Expert Judgment: A Practical Guide. London: Academic Press Ltd.; 1991.

18. Phillips LD. Group elicitation of probability distributions: Are many heads better than one. 1999.

19. Clemen RT, Reilly T. Making hard decisions with Decision Tools. Pacific Grove, CA: Duxbury; 2001.

20. Walls L, Quigley J. Building prior distributions to support Bayesian reliability growth modelling using expert judgement. Reliability Engineering and System Safety. 2001;74(2):117-28.

21. Speirs-Bridge A, Fidler F, McBride M, Flander L, Cumming G, Burgman M. Reducing overconfidence in the interval judgments of experts. Risk Analysis. 2010;30(3):512-23.

22. Anadon LD, Nemet G, Verdolini E. The future costs of nuclear power using multiple expert elicitations: effects of RD&D and elicitation design. Environmental Research Letters. 2013;8(3):034020.

23. Verdolini E, Anadon LD, Lu J, Nemet GF. The effects of expert selection, elicitation design, and R&D assumptions on experts' estimates of the future costs of photovoltaics. Energy Policy. 2015(in press).

24. Victor D, Zhou D. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2014.

25. Kriegler E, Weyant J, Blanford G, Krey V, Clarke L, Edmonds J, et al. The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. Climatic Change. 2014;123(3-4):353-67.

26. Anadon LD, Bosetti V, Bunn M, Catenacci M, Lee A. Expert Judgments about RD&D and the Future of Nuclear Energy. Environmental Science \& Technology. 2012;46(21):11497-504.

27. Henrion M, Fischhoff B. Assessing uncertainty in physical constants. American Journal of Physics. 1986;54(9):791-8. doi: doi:http://dx.doi.org/10.1119/1.14447.

28. Montibeller G, von Winterfeldt D. Cognitive and Motivational Biases in Decision and Risk Analysis. Risk Analysis. 2015;35(7):1230-51. doi: 10.1111/risa.12360.

29. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185(4157):1124-31.

30. O'Hagan A, C.E. Buck, A. Daneshkhan, J.R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakey, and T. Rakow Uncertain Judgments: Eliciting Experts' Probabilities: John Wiley & Sons, Ltd.; 2006.

31. Bosetti V, Catenacci M, Fiorese G, Verdolini E. The future prospect of PV and CSP solar technologies: An expert elicitation survey. Energy Policy. 2012;49:308-17.

32. Zickfeld K, Morgan MG, Frame DJ, Keith DW. Expert judgments about transient climate response to alternative future trajectories of radiative forcing. Proceedings of the National Academy of Sciences. 2010;107(28):12451-6.

33. Broeze KA, Brent C. Opmeer, Fulco van der Veen, Patrick M. Bossuyt, Siladitya Bhattacharya, and Ben W.J. Mol Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. Human Reproduction Update. 2010;Vol.16(No.6):pp. 561–7.

34. Riley RD. Commentary: Like it and lump it? Meta-analysis using individual participant data. International Journal of Epidemiology. 2010;Vol.39:1359–61.

35. Verdolini E, Galeotti M. At home and abroad: An empirical analysis of innovation and diffusion in energy technologies. Journal of Environmental Economics and Management. 2011;61(2):119-34. doi: 10.1016/j.jeem.2010.08.004.

36. Weyant JP. Accelerating the development and diffusion of new energy technologies: Beyond the "valley of death". Energy Economics. 2011;33(4):674-82. doi: 10.1016/j.eneco.2010.08.008.

37. Baker E, Bosetti V, Anadon LD. Special issue on defining robust energy {R\&D} portfolios. Energy Policy. 2015;80(0):215-8.

38. Bosetti V, Marangoni G, Borgonovo E, Diaz Anadon L, Barron R, McJeon HC, et al. Sensitivity to energy technology costs: A multi-model comparison analysis. Energy Policy. 2015;80:244-63. doi: http://dx.doi.org/10.1016/j.enpol.2014.12.012.

39. Gillingham K, Deng H, Wiser RH, Darghouth N, Nemet G, Barbose GL, et al. Deconstructing Solar Photovoltaic Pricing: The Role of Market Structure, Technology, and Policy. The Energy Journal. 2016;37(3):231-50.

40. Koomey J, Hultman NE. A reactor-level analysis of busbar costs for {U.S.} nuclear plants, 1970-2005. Energy Policy. 2007;35(11):5630-42. PubMed PMID: ISI:000250545300039.

41. Finkenrath M. Carbon Dioxide Capture from Power Generation – Status of Cost and Performance. Chemical Engineering \& Technology. 2012;35(3):482-8. doi: 10.1002/ceat.201100444.

42. Goldberg PK, Verboven F. Market integration and convergence to the Law of One Price: evidence from the European car market. Journal of International Economics. 2005;65(1):49-73.

43. Seel J, Barbose GL, Wiser RH. An analysis of residential PV system price differences between the United States and Germany. Energy Policy. 2014;69:216-26. doi: http://dx.doi.org/10.1016/j.enpol.2014.02.022.

44. Baker E, Chon H, Keisler J. Advanced Solar {R\&D}: Applying Expert Elicitations to Inform Climate Policy. Energy Economics. 2007;in press.

45. Abdulla A, Azevedo IL, Morgan MG. Expert assessments of the cost of light water small modular reactors. Proceedings of the National Academy of Sciences. 2013. doi: 10.1073/pnas.1300195110.

46. Baker E, Chon H, Keisler J. Advanced solar R&D: Combining economic analysis with expert elicitations to inform climate policy. Energy Economics. 2009;31(Supplement 1):S37-S49.

47. Anadon LD, Bunn M, Chan G, Chan M, Jones C, Kempener R, et al. Transforming U.S. Energy Innovation. Report. Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, 2011.

48. Curtright AE, Morgan MG, Keith DW. Expert Assessments of Future Photovoltaic Technologies. Environmental Science \& Technology. 2008;42(24):9031-8. PubMed PMID: ISI:000261678800009.

49. Fiorese G, Catenacci M, Verdolini E, Bosetti V. Advanced biofuels: Future perspectives from an expert elicitation survey. Energy Policy. 2013;56:293-311. doi: http://dx.doi.org/10.1016/j.enpol.2012.12.061.

50. Baker E, Keisler JM. Cellulosic biofuels: Expert views on prospects for advancement. Energy. 2011;36(1):595-605. doi: 10.1016/j.energy.2010.09.058.

51. Baker E, Chon H, Keisler J. Carbon capture and storage: combining economic analysis with expert elicitations to inform climate policy. Climatic Change. 2009;96(3):379-408.

52. Chan G, Anadon LD, Chan M, Lee A. Expert elicitation of cost, performance, and {RD\&D} budgets for coal power with {CCS}. Energy Procedia. 2011;4:2685-92. doi: 10.1016/j.egypro.2011.02.169.

FIGURES

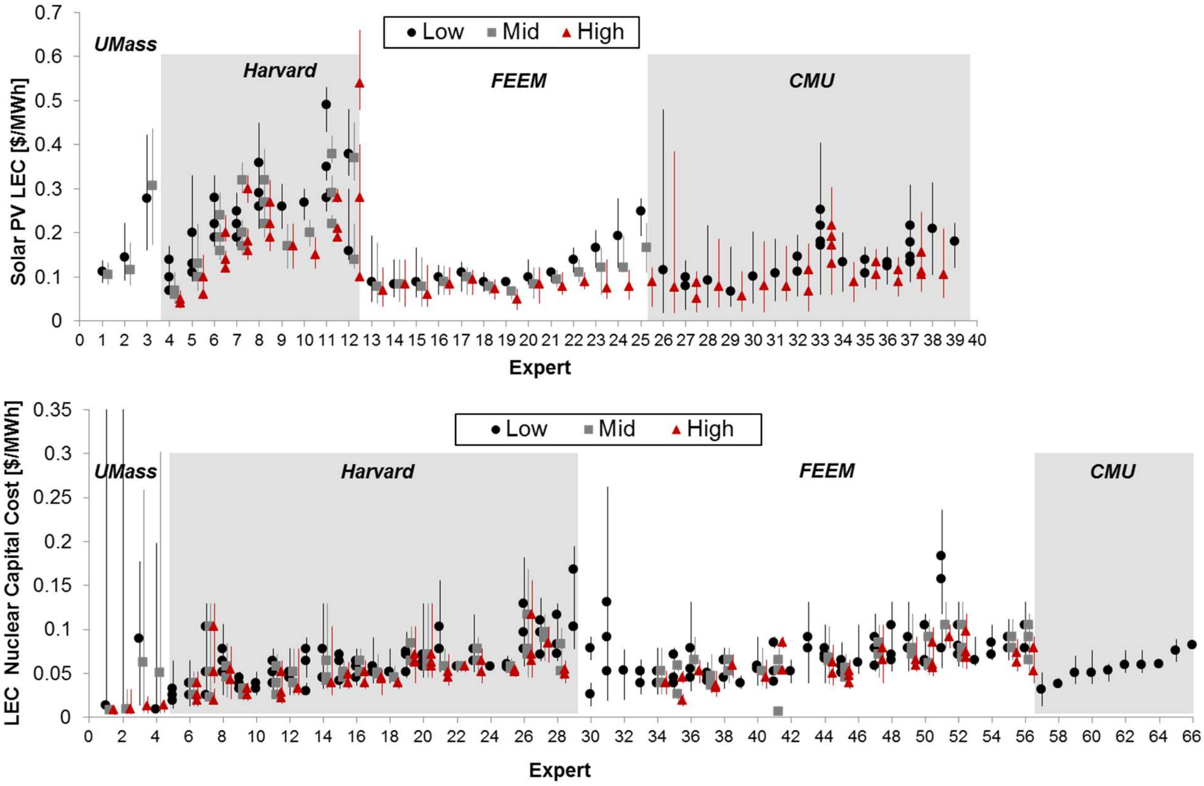

Figure 1. Levelized cost estimates for 2030 from each expert for solar PV (top panel) and nuclear power (bottom panel). Circles correspond to the P50 for the "low" public R&D budget scenario, squares for mid R&D, and triangles for high R&D. Colored lines from markers extend to P10 and P90 levels: black for Low, grey for Mid, and red for High R&D levels. Experts are grouped by study and sorted by P10 response for low R&D. All costs are in $/MWh. The grey and white backgrounds separate surveys, with the name of the group conducting each study shown in black. Two UMass nuclear expert provided low R&D values that are not shown, since we cut the y-axis at 0.35 to facilitate the inspection: expert #1 has a 90th percentile estimate at 0.80 and expert #2 has 50th and 90th percentiles of 0.78 and 1.65 $/MWh, respectively.

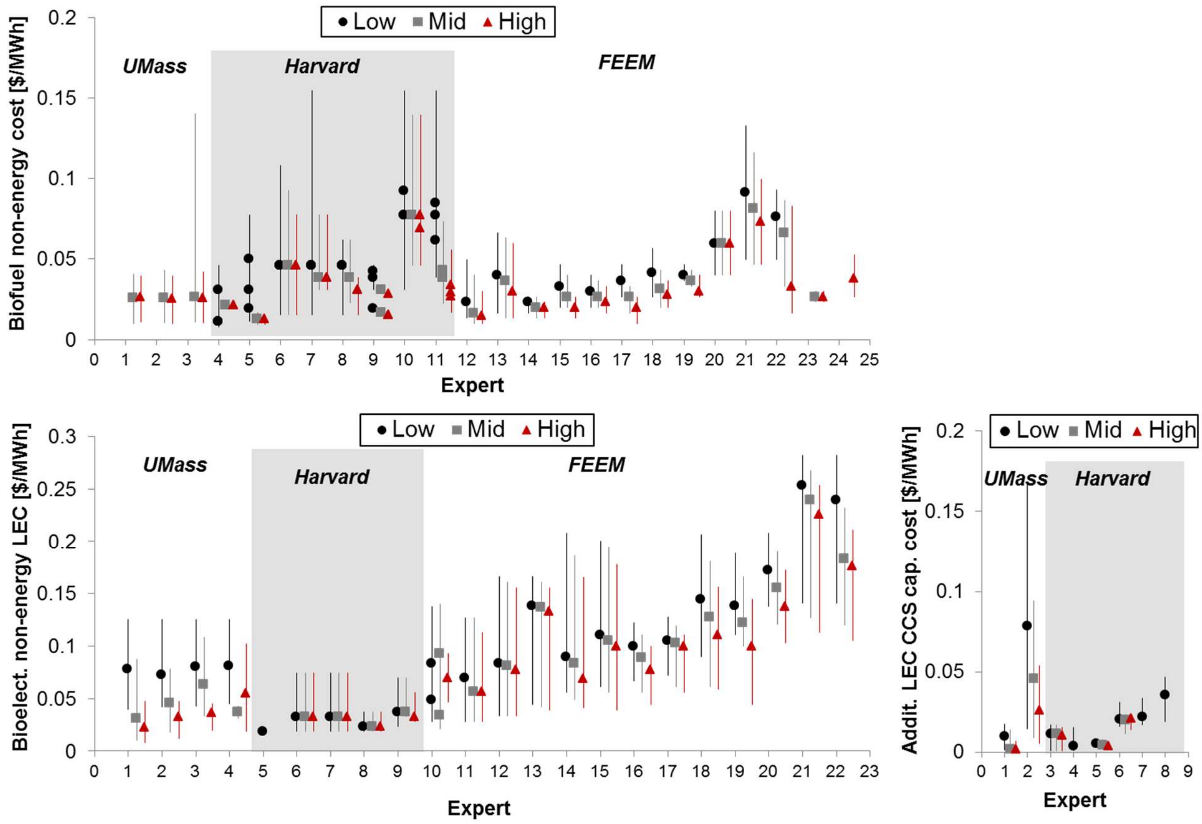

Figure 2. Levelized cost estimates for 2030 from each expert for biofuel non-energy cost (top-panel) and for bioelectricity non-energy cost and additional CCS levelized capital cost (both on the bottom panel). Circles correspond to the P50 for the "low" public R&D budget scenario, squares for mid R&D, and triangles for high R&D. Colored lines from markers extend to P10 and P90 levels: black for Low, grey for Mid, and red for High R&D levels. Experts are grouped by study and sorted by P10 response for low R&D. All costs are in $/MWh. The grey and white backgrounds separate surveys, with the name of the group conducting that study shown in black.

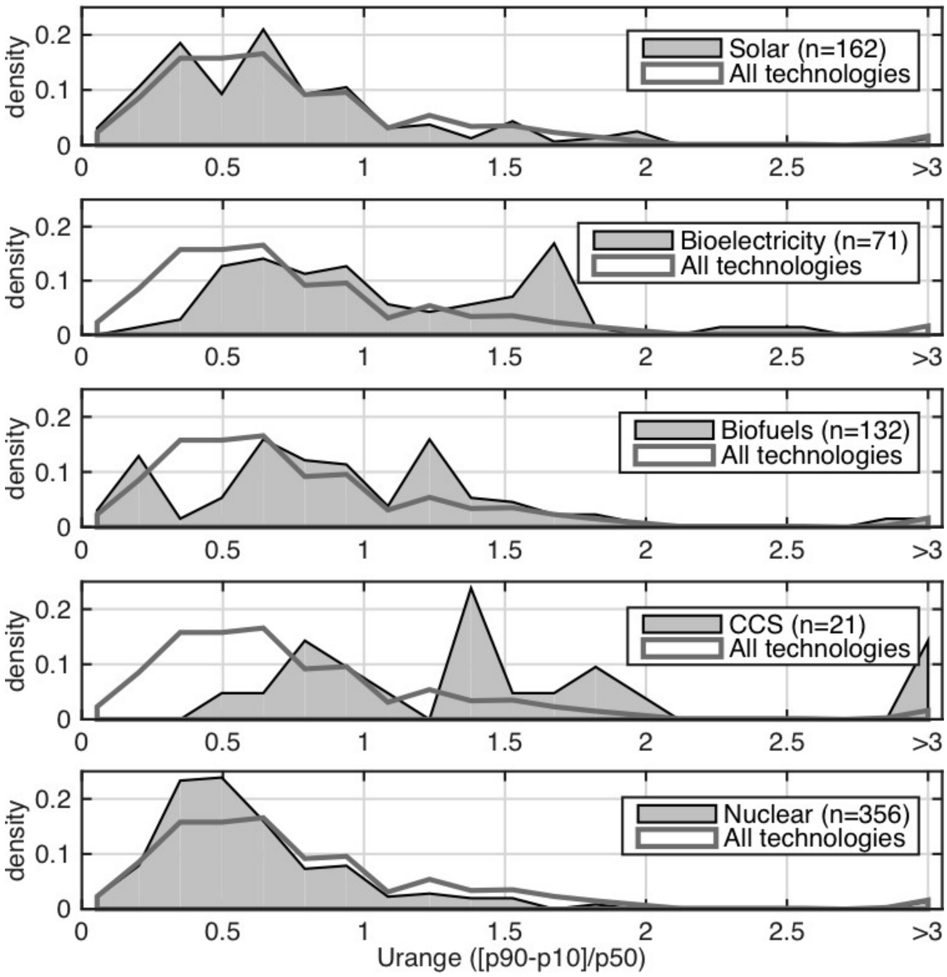

Figure 3. Distributions of uncertainty range (Urange) for all elicitations and R&D levels: pooling all five technologies (line) and for each technology individually (gray area).

TABLES

Table I. Characteristics of expert elicitation studies used.

Study code	Group(ref)	Experts	Obs	Year of elicitation	In-person	Bins for study's R&D scenarios			
						Drop	Low	Mid	High
1. Nuclear									
11	UMass[44]	4	12	2007	YES		Low	Base	High
12	Harvard[26]	25	162	2010	NO	0.5X	BAU	Rec	10x
13	FEEM[26]	30	172	2011	NO	0.5X	BAU	Rec	10x
14	CMU[45]	12	10	2011	YES		Status quo	--	--
2. Solar									
21	UMass[46]	3	6	2007	YES		Low	Mid	--
22	Harvard[47]	9	69	2010	NO	0.5X	BAU	Rec	10x
23	FEEM[31]	13	39	2011	YES		BAU	1.5x	2x
24	CMU[48]	18	48	2008	YES	10x & Dep.	Status quo	--	10x
3. Bioelectricity									
31	UMass[Unp.]	4	12	2007	YES		Low	Mid	High
32	Harvard[47]	7	21	2010	NO	0.5X	BAU	Rec	10x rec
33	FEEM[49]	16	38	2011	YES		Low	1.5x	2x
4. Biofuel									
41	UMass[50]	3	6	2008	YES			Mid	High
42	Harvard[47]	8	90	2010	NO	0.5X	Low	Rec	10x
43	FEEM[49]	15	36	2011	YES		Low	1.5x	2x
5. Carbon capture									
51	UMass[51]	3	6	2007	YES		Low	Mid	High
52	Harvard[52]	8	15	2010	NO	0.5X	BAU	Rec	10x
Total		178	742						

Table notes: BAU = business as usual; Rec = recommended; Dep= deployment; Unp.=unpublished.

Table II. Descriptive statistics for dependent (Urange, P10_LEC) and independent variables. Variables from RD_hi to bottom of table are binary.

Variable	Obs	Mean	Std. Dev.	Min	Max
Urange	742	0.869	2.186	0.054	57.54
P10_LEC	742	0.058	0.057	0	0.480
RD	694	3,549	8,629	13	80,000
RD_hi	742	0	0	0	1
RD_mid	742	0.255	0.436	0	1
RD_lo	742	0.381	0.486	0	1
Bioelec.	742	0.096	0.294	0	1
Biofuel	742	0.178	0.383	0	1
Nuclear	742	0.480	0.500	0	1
Solar	742	0.218	0.413	0	1
CCS	742	0.028	0.166	0	1
academia	742	0.330	0.471	0	1
private	742	0.395	0.489	0	1
public	742	0.275	0.447	0	1
EU	742	0.380	0.486	0	1
Inperson	742	0.287	0.453	0	1

Table III. Factors affecting the uncertainty range, Y=ln(Urange).

Y=ln(Urange)	(1) pooled	(2) pooled	(3) Solar	(4) Nuclear	(5) Bioelec.	(6) Biofuel	(7) CCS
lnperson	0.288***	0.287***	0.317***	0.455***	0.0784	0.333*	0.627***
	[0.00010]	[0.00075]	[1.99e-06]	[0.00326]	[0.613]	[0.0650]	[0.00382]
academia	0.110**	0.109**	0.0115	0.189**	0.00537	0.0357	0.00387
	[0.0302]	[0.0394]	[0.890]	[0.0131]	[0.968]	[0.816]	[0.992]
public	0.0626	0.0771*	-0.0673	0.156***	0.105	0.00809	-0.173
	[0.130]	[0.0838]	[0.286]	[0.00578]	[0.402]	[0.952]	[0.646]
EU	-0.129***	-0.132***	-0.0878	-0.0374	-0.0993	-0.404**	
	[0.00170]	[0.00416]	[0.254]	[0.416]	[0.388]	[0.0280]	
RD_hi	0.00271		0.0705***	-0.00858	0.0502**	-0.138**	0.140
	[0.871]		[0.00133]	[0.641]	[0.0334]	[0.0244]	[0.488]
RD_mid	0.00790		0.0339*	0.000876	0.0463	-0.0931*	0.217
	[0.623]		[0.0597]	[0.957]	[0.175]	[0.0946]	[0.441]
ln(RD)		-3.390					
		[0.227]					
ln(RDsq)		1.690					
		[0.228]					
Nuclear	0.156**	0.172**					
	[0.0330]	[0.0110]					
Bioelec.	0.170***	0.193***					
	[0.00891]	[0.00539]					
Biofuel	0.164**	0.166**					
	[0.0215]	[0.0237]					
CCS	0.493***	0.529***					
	[0.00023]	[0.00013]					
Obs.	678	694	162	322	66	110	18
# of experts by subtech	301	276	71	159	23	40	8
# Clusters	160	146	39	66	23	24	8
R2 overall	0.228	0.223	0.319	0.174	0.0395	0.128	0.378
R2 within	0.000796	0.0348	0.166	0.00362	0.0606	0.154	0.186

Clustered p-values in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table IV. Factors affecting the "best" outcome, Y=ln(P10), for individual technologies.

Y=ln(P10) VARIABLES	(1) Solar	(2) Nuclear	(3) Bioelectricity	(4) Biofuel	(5) CCS
Inperson	-0.0845***	-0.0202***	0.00709	-0.0163***	-0.00359
	[6.09e-07]	[4.00e-06]	[0.518]	[0.000342]	[0.617]
academia	0.000778	-0.0216***	-0.00691	0.00995**	-0.0102*
	[0.959]	[2.83e-07]	[0.607]	[0.0149]	[0.0714]
public	0.0194	-0.0118**	-0.0290**	0.0137**	-0.00131
	[0.213]	[0.0101]	[0.0409]	[0.0397]	[0.871]
EU	-0.0121	0.00349	0.0437***	0.0128***	
	[0.209]	[0.385]	[0.00103]	[0.000357]	
RD_high	-0.0316***	-0.00961***	-0.0169***	-0.00465*	-0.00265
	[1.63e-08]	[0]	[5.45e-06]	[0.0944]	[0.151]
RD_mid	-0.0140***	-0.00478***	-0.00808***	-0.00169	-0.00221
	[0.000188]	[0.000221]	[0.000361]	[0.508]	[0.107]
Observations	162	322	66	110	18
Number of experts, by subtech	71	159	23	40	8
Nr Clusters	39	66	23	24	8
R2 overall	0.419	0.284	0.477	0.240	0.356
R2 within	0.371	0.372	0.476	0.119	0.365

Clustered pvalues in brackets

*** p<0.01, ** p<0.05, * p<0.1

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S-1. Individual elicitation results for solar PV (top-panel) and for nuclear power (bottom panel).

Figure S-2. Individual elicitation results for biofuel non-energy cost (top-panel) and for bioelectricity non energy cost and additional CCS levelized capital cost (both on the bottom panel).

Figure S-3. Distribution of experts' Urange estimates for all observations.

Table S-1. Assignment of study-specific R&D levels to standardized R&D bins.

Table S-2. Definitions of technologies and sub-technologies included in the elicitations.

Table S-3. Definitions of variables used.

Table S-4. Estimates for models of Y = ln(P10), with continuous R&D.

Table S-5. Estimates of models for Y = ln(P50).

Table S-6. Estimates of models for Y = ln(P90).

Table S-7. Correlation matrix of covariates.