NOTA DI LAVORO

1.2014

Facing the Experts: Survey Mode and Expert Elicitation

By Erin Baker, University of Massachusetts at Amherst

Valentina Bosetti, Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università Bocconi

Karen E. Jenni, Insight Decisions LLC

Elena Claire Ricci, Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università degli Studi di Milano

Climate Change and Sustainable Development
Series Editor: Carlo Carraro

Facing the Experts: Survey Mode and Expert Elicitation

By Erin Baker, University of Massachusetts at Amherst
Valentina Bosetti, Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università Bocconi
Karen E. Jenni, Insight Decisions LLC
Elena Claire Ricci, Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università degli Studi di Milano

Summary

In this paper we compare the results of two different expert elicitation methods: in-person interviews and a self-administered web-based survey. Traditional expert elicitation has been done face to face, with an elicitor meeting with an expert for a few hours to several days, depending on the complexity of the analysis. Recently, however, some groups have been using other methods to solicit expert judgments, including self-administered surveys (written, emailed, and web-based), and the use of interactive web tools to facilitate interactions during an elicitation. These elicitations require fewer resources from the assessment team than in-person interviews, and often allow participating experts to provide input on their own schedules, perhaps with additional time to think about their responses. Thus they open up the possibility of using expert elicitation to obtain inputs relevant to a broader set of decisions. To our knowledge, these newer survey-based methods have not been rigorously evaluated for efficacy. We find, much like the results in the literature on different survey modes, different results from two different modes we examined, but no clear indication of which method might be preferred. We suggest future work including some controlled, lab-based experiments and real EEs well designed to avoid sample selection biases and specifically targeted to capture survey mode effects. Such studies would help us determine whether and when the different survey modes are most effective.

Keywords: Expert Elicitation, Interviews, Web Surveys, Carbon Capture and Storage

JEL Classification: D81, O32, Q55, Q40

This project was partially supported by NSF under award No. SMA-0960993.

Address for correspondence:

Elena Claire Ricci
Fondazione Eni Enrico Mattei
Corso Magenta 63
20123 Milan
Italy
E-mail: elena.ricci@feem.it

Facing the experts: survey mode and expert elicitation

Erin Baker[1], Valentina Bosetti[2], Karen E Jenni[3], Elena Claire Ricci[4]

[1] University of Massachusetts at Amherst
[2] Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università Bocconi
[3] Insight Decisions LLC
[4] Fondazione Eni Enrico Mattei, Centro Euro-Mediterraneo per i Cambiamenti Climatici and Università degli Studi di Milano

ABSTRACT

In this paper we compare the results of two different expert elicitation methods: in-person interviews and a self-administered web-based survey. Traditional expert elicitation has been done face to face, with an elicitor meeting with an expert for a few hours to several days, depending on the complexity of the analysis. Recently, however, some groups have been using other methods to solicit expert judgments, including self-administered surveys (written, emailed, and web-based), and the use of interactive web tools to facilitate interactions during an elicitation. These elicitations require fewer resources from the assessment team than in-person interviews, and often allow participating experts to provide input on their own schedules, perhaps with additional time to think about their responses. Thus they open up the possibility of using expert elicitation to obtain inputs relevant to a broader set of decisions. To our knowledge, these newer survey-based methods have not been rigorously evaluated for efficacy. We find, much like the results in the literature on different survey modes, different results from two different modes we examined, but no clear indication of which method might be preferred. We suggest future work including some controlled, lab-based experiments and real EEs well designed to avoid sample selection biases and specifically targeted to capture survey mode effects. Such studies would help us determine whether and when the different survey modes are most effective.

Keywords: expert elicitation, interviews, web surveys, carbon capture and storage

JEL codes: D81, O32, Q55, Q40

1. INTRODUCTION

This paper is largely motivated by interest in applying science to science policy. A National Academy Study suggested that sound government R&D policy related to technology development should consider the *likelihood of success* and the *impacts of success*, along with the total cost of a program, when making funding decisions.[1] Generally, estimating the likelihood of success for different investment decisions requires expert judgments, and obtaining a sufficient number and quality of those judgments is quite challenging. The US Department of Energy, for example, has 10 major technology categories, each with a number of sub-categories, and each subcategories may have multiple technologies of interest. There are hundreds of possible technologies that would need to be characterized to do a full portfolio analysis of government-supported energy programs in the US. While a number of studies have a made a start,[2-6] it is a daunting prospect to collect expert input on hundreds of disparate energy technologies, and it is natural to look for less resource- and time-intensive approaches for obtaining that input than in-depth, face-to-face expert elicitations.

One approach gaining popularity is to streamline the elicitations by using web-based tools. However, we are aware of no studies examining the effectiveness of web-based approaches or comparing them to traditional interview-based expert elicitations. This paper describes an initial exploration of the similarities and differences between the two elicitation modes.

We first review of the literature describing traditional approaches to expert elicitations and then, since there is no literature comparing different elicitation modes, we turn to the literature comparing different modes of surveys for insights on the effect of elicitation mode. In Section 2 we describe and an exploratory analysis comparing two elicitations of an emerging energy technology, Carbon Capture & Storage (CCS). One elicitation was conducted via a traditional interview mode, and the other via a self-administered web-based survey. Both studies asked for the same type of expert assessment of the same quantities, and an attempt was made to make the questions as similar as possible, providing some degree of comparability between the two survey modes. Because these are real elicitations (rather than lab-based experiments) there is no gold standard to evaluate the "quality" or "accuracy" of the responses. Thus, we focus on comparing the ways in which results from the two modes do or do not differ, including the rankings of the technologies, the level of uncertainty present in the assessments, the absolute values assessed, and on process variables such as the number of "Don't Know" responses. Section 3 closes with a discussion of insights from the comparison and areas for future study.

1.1. What is "expert elicitation?"

"Probability encoding, the process of extracting and quantifying individual judgment about uncertain quantities…" (Spetzler and Stael von Holstein[7]).

A fundamental aspect of decision analysis is quantification of uncertainties using probability. Early in the practice of decision analysis, analysts faced the challenge of figuring out how to obtain either decision-maker or expert judgments and translate them into probabilities that could be used in a decision tree. Even earlier, issues associated with obtaining "accurate" probabilistic estimates or forecasts were recognized. For example, Brier[8] proposed a "verification formula" to improve the calibration of statistically knowledgeable experts, specifically weather forecasters. Winkler[9] compared four approaches for assessing prior distributions for Bayesian analyses, emphasizing that the sources of those distributions might not be statistically sophisticated. This recognition that those who are most knowledgeable about the quantities that a decision-maker or analyst cares about might not be particularly adept at thinking and expressing themselves in terms of probability, combined with emerging research on the various "heuristics and biases" that affect judgments under conditions of uncertainty[10] led to the development of several structured approaches or protocols for obtaining probability estimates from non-statistician experts.

One of the first robust descriptions of an elicitation protocol identified the process as "probability encoding," defined as shown in the quote above. This terminology suggests a mindset where experts are viewed as having a well-formulated judgment about the quantity of interest, and the task is simply to obtain that judgment in a way that an analyst can use it (as some statement of probability): to ask the right questions in the right way to "extract" those judgments. At the same time, others in the field explicitly acknowledged that experts are not likely to have such well-formulated judgments: "It must be stressed that the assessor has no built-in prior distribution which is there for the taking. That is, there is no 'true' prior distribution… An elicitation technique used by the statistician does not elicit a 'pure' prior distribution, but in a sense helps to draw out an assessment of a prior distribution from the prior knowledge" (Winkler[9]).

Early elicitation protocols combined these two perspectives, emphasizing the interactive nature of the elicitation process, with the analyst working together with an expert as she develops probability distributions that represent her knowledge. However, the approach and the protocols also included a tight focus on obtaining estimates of specific individual quantities of interest to the analyst and / or decision-makers. Morgan and Henrion[11] describe several such protocols, all of which include some version of five basic steps in what became known as the "SRI protocol": (1) motivate the assessment and explore possible biases, (2) structure the uncertain quantity to be assessed, (3) condition the expert to try to avoid cognitive biases as she approaches the elicitation task, (4) encode the experts judgments as probability distributions, and (5) verify the results through consistency checks. Much of the emphasis was on defining approaches and questions that could be used for the "encoding" step itself, many of which are reviewed in Hora[12].

Over time, the term "expert elicitation" has come to be interpreted more broadly, encompassing not just quantitative estimates of clearly defined uncertain values, but also including expert input on conceptual model design or selection, scenarios, and a full range of modeling choices as well as model inputs. Expert elicitation has been called:"…a structured procedure designed to gather knowledge… from individuals considered human experts in that domain. Topics of elicitations can be probability encoding, scenario development, or model selection." (DeWispelare *et al*.[13]).

Fischhoff[14] described how expert knowledge can be "elicited" for each of the key stages in an analysis; specifically, to help identify the structure of the problem and the relationships between model elements, to estimate quantitative parameters for the resulting model, and to evaluate the quality of the model representation. Recent work characterized as "expert elicitation" often encompasses all of these stages, with analysts working with experts to define the model structure and relationships before assessing (or "eliciting") input on the quantitative components of the model. Much of this work also involves more than one expert, bringing into play questions about whether and how the input from multiple experts ought to be combined (e.g., Cooke[15], SSHAC[16]).

1.1.1. How are "expert elicitations" carried out?

While there is a rich literature on expert elicitation approaches and protocols, there is less information available on the specifics of how an elicitation is carried out. Much of the literature cited above seems to assume that the elicitation process will be conducted in a face-to-face setting involving individuals or small groups. Other approaches, such as the Delphi method,[17] place more emphasis on the exchange of written materials, especially in the review and update of initial assessments. When elicitations involving multiple experts are carried out, additional steps are necessary to ensure that appropriate and comparable input is being obtained from each of those experts. Jenni and van Luik[18] reviewed several documents describing or recommending a set of steps for formal expert elicitations with multiple experts (e.g., Kotra *et al*. [19], SSHAC[16], Cooke and Gossens[20]) and concluded that all of these approaches share an emphasis on the steps listed in Table I.

Meyer and Booker[21] describe two of the key elements of conducting an elicitation as determining "the setting in which the elicitation of the expert's judgment takes place" and choosing an appropriate "means by which the data-gatherer and the expert communicate" during the elicitation process. These two elements are closely related, with some means of communication being more effective in various settings. Approaches differ in the degree of interaction each allows or promotes between experts, and between the analyst and the expert, as shown in Figure 1.

Interactive group elicitation settings and individual interviews necessarily involve face-to-face or personal interactions between the analyst and the expert, and such interactions are generally considered the "gold standard" for elicitations: "if eliciting deep problem-solving data is the goal, this is the only suitable mode of communication" (Meyer and Booker[21]). Meyer and Booker[21]discuss some of the advantages of elicitation modes using either telephone or mail (postal or electronic), primarily that they can be done with less expense and can, potentially, involve a greater number of experts. Each of these less resource-intensive approaches has disadvantages, such as inability to cover topics in depth and response rate concerns for mailed surveys. They specifically suggest that "complicated response modes that require training," which would include most probability-encoding type questions, "should not be used" in telephone or mail surveys.

Despite these concerns, it is becoming more common for researchers to conduct expert elicitation studies and other formal assessments of expert knowledge via survey-type instruments.[22-25] Several groups have developed and are testing interactive web-based tools as a method for conducting both analyst-facilitated and self-administered expert elicitations.[26-30] While there is an extensive literature on the implications of different methods for carrying out surveys of the general public, to our knowledge survey-based expert elicitation methods, and more generally the implications of survey mode, have not been rigorously evaluated.

1.2. Survey modes and their impact

The impact of survey mode, including face to face interviews, interviews over the phone, and self-administered modes such as mail or internet surveys, for general public surveys has been extensively studied. The mode has been shown to influence the quality and accuracy of the data collected, yet the overall findings on which methods yield more accurate or higher quality data are mixed. Several studies have compared the results of self-administrated surveys (via mail or internet) with the results of interviews (telephone or face-to face), especially in the areas of public health and contingent valuation. In a review of health and epidemiological questionnaires, Bowling(31) concludes that administration methods have important repercussions on survey responses, with particularly strong effects are seen between self-administration and interviews. The two main differences between these survey modes relate to the cognitive effort required from the respondent and the level of anonymity provided.

A common hypothesis is that self-administered surveys may lead to satisficing: taking cognitive shortcuts to make it easier and faster to complete the survey,(32-33) and that satisficing leads to less accurate responses. Interview-type surveys are thought to require less cognitive effort from respondents than self-administered surveys, because the respondents need only pay attention to the interviewer (no reading skills required) and can ask for clarifications. In addition, respondents may be motivated to answer the questions more thoughtfully and accurately by the presence of an interviewer.(31)

In order to test the satisficing hypothesis a number of studies have looked at (1) the response variability among questions (whether people are more prone to give similar answers to similarly structured questions), (2) the number of non-responses or "Don't Knows" (DK), and (3) time to complete the survey. Typically lower response variability and a high number of DKs are interpreted as indications that the respondents are satisficing. Table II summarizes the conflicting results of three studies comparing self-administered and interview-based surveys in terms of the evidence they find for satisficing.

Interpretation of the time taken to complete a survey is mixed. Fricker *et al.*(34) found that web respondents took less time to complete closed-ended questions than interview respondents, but took longer overall because of the time spent responding to open-ended questions. They interpret the latter finding as evidence of a higher cognitive load for self-administration. Heerwegh and Loosveldt(35) found that web respondents took less time to complete the survey, which they interpreted as evidence of satisficing (respondents were not paying as much attention). Both studies interpret their opposite findings in a way that reflects negatively on self-administered web-based surveys relative to interviews.

One area in which self-administered surveys are generally believed to yield more accurate responses is when the questions relate to issues or behaviors where there are strong social norms: the social-desirability bias may lead interview respondents to give answers that comport with behavior that is considered socially acceptable or desirable, whether those answers are accurate or note. Many, but not all, studies conclude that this bias is stronger in interviews than in self-administered surveys.(36-42) It is not clear how this question is relevant in the context of a typical expert elicitation.

Finally, there are some results that indicate web-based modes may lead to more accuracy in knowledge questions in which visual aids were considered to be particularly important.(34,43)

1.3. Possible effects of elicitation mode

Expert elicitations differ from general public surveys in important ways, but some of the factors that have been shown to be relevant in public survey response are also relevant for elicitation. For example, we expect that the interview mode poses a lower cognitive burden on the experts and promotes clarity of responses. Interview participants have the ability to ask questions and clarify the task as they go, and are provided with immediate feedback on their responses and the interpretation of those responses. Because of the interactive nature of the interview process, the analyst can be fairly confident that she knows what the expert was thinking and why specific responses were provided. These factors suggest that interviews provide more and better quality information. We note, however, in our own personal experience we have sometimes received very high quality written comments and explanations from experts participating in written surveys.

A potential advantage of web surveys of experts over interviews is that respondents have as much time as they need to develop a response, and can access research materials and talk to others if they so choose. This absence of time pressure and ability to collect information to help inform their responses suggest that for a highly motivated respondent, the quality of information may be higher in web-based elicitation than an interview-based elicitation.

In a real elicitation, "quality" of the response is difficult to judge, as there is no data against which expert responses can be compared. Thus, we focus this exploratory analysis on the key question of interest to a study sponsor or designer: Do the different modes lead to different conclusions? To explore that question we considered several different ways in which "conclusions" can be drawn from these elicitation results. Specifically, we looked at:

- Level of uncertainty expressed (the range of values)

- Ranking of the specific items being studied (carbon capture and storage technologies under different policy scenarios)
- Differences in the individual and aggregate results of the elicitations
- Qualitative differences related to the elicitation process

2. Case study: Expert interviews and web-based expert survey of carbon capture technologies

2.1. Study background

We have recently completed an expert elicitation study to explore expert opinion about the energy penalty (EP), the energy required to capture and compress CO2 from power plants, for multiple carbon capture technologies given different policy scenarios[44]. Carbon capture and storage is an emerging technology aimed at capturing the emissions of CO_2 before they are released from fossil or biomass-based power plants, and storing the captured CO_2 so that it is not released into the atmosphere. It is considered an important part of a comprehensive climate strategy, but it is a young and still-evolving technology. Questions exist about the storage, the capital cost, and the operating costs for CCS. We focused on this last question, specifically addressing the EP, an element of operating cost that may be reduced by different government policies, including increased spending on R&D. We asked experts for their evaluation of the energy penalty associated with six carbon capture technologies[5] in 2025, under three different policy scenarios.[6] While designing the initial interview-based elicitations we identified an opportunity to conduct a parallel set of elicitations using the web-based tools being

[5] The technologies taken into account are: absorption/solvents, adsorption, membranes and ionic liquids, other post-combustion technologies such as enzymes or cryogenics, pre-combustion, and oxyfuel.

[6] A "no further government R&D" scenario; a carbon-policy/carbon-pricing scenario; and an increased R&D investment scenario.

developed and utilized at FEEM (http://www.icarus-project.org). It is important to note that we did not set out to design a rigorous, controlled study of the differences between these two elicitation modes. We did endeavor to make the two elicitations as similar as we could – asking the same questions about the same quantities, and providing much of the same background material to the expert participants in each study. We view this as an exploratory comparison of two elicitation modes: traditional interview-based elicitation (Interview) and self-administered survey-based elicitations over the web (Web).

2.1.1. Elicitation steps for each study

Below we summarize how each of the seven basic steps from Table I were implemented for the two studies.

Define the objectives of the study and determine whether expert elicitation is necessary. The primary purpose of this study was to improve our understanding of how different policy "levers" might affect research and development outcomes related to reducing the costs of CCS technologies, as part of a larger effort to inform science policy both as it relates to the specific technologies being considered, and as it relates to the effectiveness of different policy levers. While some of the cost related factors can be modeled with engineering cost models and other established approaches,[45] there are no data or models to support estimates of how CCS technologies will evolve under different policies. Those estimates require expert judgment.

Select the experts. For the Interview elicitations, we identified possible participants through a review of the literature and through discussions with several technical advisors to the project. The study focus was on US policies, so we focused on US experts, but we also included two experts from the EU because of their breadth of expertise. Potential experts were recruited by email and phone. A total of 15 experts participated in the elicitation interviews, 13 in person and 2 by phone.

For the Web elicitations, we identified a large number of experts in the field of CCS, identifying authors from leading journals and from prominent studies such as the IPCC. We identified 236 experts, who were recruited over email. Forty-one experts responded to the initial invitation, and 12 experts[7] completed the web-based survey, 10 from the EU, 1 from the US, and 1 from South Africa.

Structure the assessment, identify and clarify the assessment issues, develop any assessment protocols necessary for the elicitation method. For both elicitation modes, the structure of the assessment was the same: for each technology, experts were asked about the potential range of EP for that technology in 2025, based on three scenarios related to government investment in R&D and carbon policies. Specifically, they were asked to provide estimates of the 5th, 50th, and 95th percentiles of a distribution describing their beliefs about the future energy penalties. A protocol for the interviews was developed, and specific wording for assessment questions in the web survey was also developed.

Develop and provide the experts with background and training about assessment tasks, cognitive biases, and probability encoding concepts. Experts in both modes were provided with the typical background materials: a summary of the goals of the project and the elicitation, a description of technologies and policy scenarios to be discussed, and an introduction to the specific value (the EP) that was to be elicited. Two additional background documents were provided to the interview participants: a fairly extensive set of sensitivity analyses that included information on some of the technical factors affecting the EP for each technology, and a short description of what to expect in the interviews. The interview elicitations also began with a presentation and discussion of probability encoding and the relevant cognitive biases. Web participants had available at the initiation of the survey (and anytime during the survey) a table depicting the various technologies included in the survey, a description of the scenarios, and several different metrics that can be used for summarizing the EP.

[7] More precisely, 10 single experts and a team of two experts.

Conduct the elicitation itself and provide feedback about their elicitation results and their implications. Interview length ranged from 2 hours to 8 hours, allowing the experts to evaluate anywhere from 1 to 7 different technologies. Each expert addressed only those technologies for which he felt comfortable making assessments. The interview protocol was not rigid, and the assessment order, specific questions asked, and even the tools used to assist the expert with understanding the assessment tasks evolved over time and were tailored to each expert. Each expert talked through their assessments, explaining their thinking as they went, and these qualitative insights were an important aspect of the elicitation results not available in the web survey. As part of the interview, experts were shown their assessment results in graphical form, and prompted to think about and discuss the different results they were projecting under the different scenarios.

Early in the interviews it became clear that in some cases the participating expert simply felt that a technology would not be viable under one or more scenarios, and the original set of assessment questions did not accommodate this type of judgment. This led to a change in the interview protocol, so that experts made an explicit estimate of the probability of viability and then a conditional assessment of the EP for that technology-scenario pair assuming it was viable. Since the level of interactivity in the web-based survey was necessarily lower, our ability to address this possibility was limited to post-elicitation discussions. In these follow-up discussions, only one web participant gave us a probability of non-viability for any technology.

In the web survey, we tried to provide as much flexibility as we could for the respondents, as well as providing opportunities for them to describe their reasoning. Respondents were able to follow the order in which the questions were posed or could pick technologies in any order, depending on their preferences and knowledge. This option was meant to encourage the participant to respond for those technologies for which the expert felt more confident. The survey provided real-time feedback, with each expert's answers displayed graphically in real time so that he could see some of the implications of those assessments. An example of the format of the questions for one technology and of the type of interactive graphical displays is reported in Figure 2. Many of the respondents provided brief comments in the comment boxes provided on the web form.

Following each interview, we prepared a written summary of the discussion and the assessment results for review and comment by the participating expert. After the web surveys had been completed and the data were being analyzed, several questions arose about the responses and how experts were interpreting questions. All of the web participants were re-contacted, most by phone and two via email, to discuss their assessment results, get their comments, and, in some cases, confirm our interpretation of their answers.

*Use the individual expert's inputs to create an aggregate assessment of the quantities of interest, if desire*d, *and document the results of the assessment*. The interview results have been described in Jenni *et al.*[44]. This paper shows the aggregated assessment results for both elicitation modes and documents both sets of results.

2.2. Exploratory analysis results

In the first step of the analysis, we estimated the probability distributions over EP for the different carbon capture technologies in 2025 for each expert by fitting distributions to their assessed 5^{th}, 50^{th} and 95^{th} percentiles. For each survey method, we aggregated the distributions of each expert using a linear opinion pool approach with equal weights.[46] Aggregated distributions such as these are examples of a commonly-used output of expert elicitation studies, and are the first set of results explored.

2. 2.1. Level of uncertainty expressed

In descriptions of expert elicitations, the importance of providing proper "training" in probability encoding and making experts aware of common biases that may affect their judgments is emphasized (e.g., Table II in Cardenas *et al.*[47]). It is commonly believed that such training helps experts provide more realistic and less biased estimates of their uncertainty. In particular, analysts are often at least as interested in evaluating the tails of distributions as they are in estimating a mean or midpoint for the uncertain quantities of interest. For example, when assessing the possibility of a breakthrough in a technology, it is crucial to make sure that experts are considering a set of future states of the world and events as large as possible. In-person interviews allow the interviewer to continuously add elements and factors that might help experts include a broader range of factors that could possibly lead to unexpected outcomes. Thus, we expected that the interview mode might lead to greater expressed uncertainty in each single expert answer and in the aggregate distributions than the web-based survey mode.

Figure 3 presents our results by depicting the aggregated distributions for all technologies, the three scenarios, and both elicitation modes. Results are presented as cumulative distribution functions, with the maximum probability shown corresponding to the aggregated estimate of the probability that the technology will be viable. Clearly visible in these graphs is the most dramatic difference between the elicitation results in the two modes: experts who were interviewed expressed more uncertainty about the viability of technologies than those in the web survey. The figure also suggests that interviewed experts expressed a wider range of uncertainty in the performance of individual technologies, and greater differences between technologies, especially at the median and higher percentiles.

Three of the six carbon capture technologies were assessed by all experts in both modes as being technically viable under all scenarios, making it straightforward to compare the results of the two survey methods in statistical models. Those technologies (absorption, pre-combustion, oxyfuel) are the most mature capture options and thus might be considered those with the least uncertainty. While we recognize that our very small sample sizes limit our ability to detect statistically significant differences in responses, we developed a simple linear regression model to look for such differences. The first model explores the potential effects of technology, scenario, and survey mode on the coefficient of variation (CoV, the standard deviation divided by the mean) of the fitted distributions for the EP value provided by each single expert:

$$CoV(EP)=\alpha+\beta_{tech}\cdot technology+\beta_{scen}\cdot scenario+\beta_{surv}\cdot survey+\varepsilon\,, \tag{1}$$

Here α is the coefficient of variation of EP in the reference situation;[8] ε is an error term. This is an additive three-way ANOVA model where we controlled for three main variables that may influence the answer of the respondents. *Technology* and *scenario* are multi-level factors identifying the type of carbon capture technology or the policy scenario; *survey* refers to the type of survey. Table III provides an overview of explanatory variables considered in the current and subsequent models, and identifies the reference situation).

We also tested a second model including a random effect to control for individual expert effects. Results for both models are reported in Table IV. In both cases Scenario 3 has a significant effect. This means that when moving from the no further government R&D scenario to the increased R&D investment scenario we should expect a greater relative uncertainty or variation. For Model 2, the differences between individual experts were also significant (via an asymptotic likelihood ratio test, p-value lower than 0.0001), and Scenario 2 also had a significant effect.

---

[8] The symbol $\beta_{tech}\cdot$technology represents a categorical variable modeled by a set of dummies, rather than a discrete variable from 1 to 3.

We would have expected to see larger uncertainty expressed in the interview elicitation mode, given the explicit discussion of cognitive biases and the opportunity for the elicitor to prompt the expert to think broadly about possible outcomes. This difference is not seen for the three technologies above, which could suggest either that the effect of the training is less than commonly expected, or that the training is less important or has less impact for more mature technologies.

2.1.2. Ranking of technologies

The aggregated distributions can be used to rank technologies, and for some studies this ranking may be the most relevant result. Depending on the study purpose, different ranking indexes may be appropriate. Figure 4 shows a comparison of the rankings of technologies for each scenario based on the 5^{th}, 50^{th}, and 95^{th} percentiles of the aggregated distributions. While in some cases the most relevant element for policy makers or analyst may be the central values - for example, when elicitations are used to calibrate a model using the median or mean of a distribution -, the rankings by 5^{th} and 95^{th} percentiles might be of interest if the key study questions have to do with the likelihood of reaching very low or very high values, rather than the central tendency.

Overall, in the interview results the highest ranked technologies are the more mature ones (pre-combustion, oxyfuel and absorption), for all scenarios and all percentiles except the best possible outcome (5th percentile) where “other-post combustion” technologies become more promising. If instead we look at the web results, we find that pre-combustion always occupies the top position of the rankings; oxyfuel does well only under the no further government R&D policy scenario; and absorption is generally lower ranked.

There is little to no difference in the ranking of technologies by the 95^{th} percentile between surveys or scenarios, suggesting a consistent view of how the technologies will perform relative to each other under a “worst case” outcome for each technology. We note that the data shows smaller differences between technologies in the web survey than in the interviews, and in some cases the values of these percentiles are very similar. Looking at the ranking overemphasizes the differences between the assessments of technologies, particularly for the web results.

2.1.3. Differences in absolute values of the assessed EP

It is not clear a priori whether or how the elicitation mode might impact the various percentile estimates of the quantity being studied. To evaluate this, we explored whether the percentiles reported by experts are affected by the survey mode by means of a regression model where the reported percentiles of the single expert EP distributions are the dependent variables, and the technology, scenario, and survey mode are possible explanatory variables:

$$EP=\alpha+\beta_{tech}\cdot technology+\beta_{scen}\cdot scenario+\beta_{surv}\cdot survey+\varepsilon\,. \qquad (2)$$

α is EP (5th, 50th, or 95th percentile) in the reference situation, defined as in model 1 and Table III,. As before, we considered only the three most mature technologies; those which all experts estimated would be viable under all conditions.

Results, reported in Table V, indicate that all the explanatory factors are significant, including elicitation mode. The predicted median value of the EP for the reference situation is about 0.20 and is significantly different from zero ($p < 0.001$). The effect of the survey mode is significant and positive for all percentiles, meaning that the interviewed experts are consistently less-optimistic than the web respondents about the future performance of carbon capture technologies in terms of induced energy penalty. Indeed, the predicted median of the future *EP* for the interviewed experts is 0.04 higher than the web-survey ones, hence survey mode is influencing the predicted median as much or more than as the policy scenario does. We also see that experts consider that EP would be reduced by 0.02 given a worldwide carbon policy or by 0.04 given increased government R&D investment. The effect of the scenario is similar across all percentiles.[9]

We were somewhat surprised by these results, having anticipated consensus across the two survey modes on the 50th percentile for these mature technologies. An alternative explanation of the difference in the results of the two survey modes (rather than a survey effect) might be different perspectives deriving from the nationality of experts rather than the survey mode itself. We added a country factor to the regression model to take into account the nation in which the experts work, to test this alternative:

$$EP=\alpha+\beta_{tech}\cdot technology+\beta_{scen}\cdot scenario+\beta_{surv}\cdot survey+\beta_{country}\cdot country+\varepsilon\,. \qquad (3)$$

Table VI shows that once the country factor is introduced the survey factor becomes less significant. US experts seem to be more pessimistic about the future prospects of carbon capture technologies in terms of energy penalty compared to European ones. While this conforms to our personal experiences and perspectives, these results must be treated cautiously as there is a near-perfect confound: almost all the interviewed experts are from the USA and almost all of the web-respondents are European.

2.1.4. Process observations.

From the literature review above on surveys, we might expect other differences between the two types of elicitations. For example, we might expect fewer DK answers in the in-person interviews – where a DK is usually interpreted as a sign of less commitment. We might expect differences in the

---

[9] We note that the difference between the beta coefficients for the two scenarios is not statistically significant, suggesting that we could create a more parsimonious model by treating scenarios 2 and 3 as a single scenario, compared with scenario 1.

amount of time spent by the experts on each question, although it is not clear in which mode we would expect more time to be spent. In the interview elicitations, the participants were explicitly told that they could provide assessments for any technology they felt they had expertise to discuss, and were able to decline to provide assessment for technologies they knew less about. Only one expert assessed all technologies, and 6 out of 15 assessed only two technologies. In the web survey, one expert assessed all technologies, one expert expressed judgments just for one technology, and most experts assessed most technologies. Interpretation of this difference is not clear: the fact that web respondents evaluated more technologies could be taken as a sign of commitment on the part of the web-respondents, or, it could reflect the fact that web respondents had the chance to access additional information for formulating their estimates and to complete the survey over whatever time period they chose, or it could indicate that the web respondents were over-confident in their abilities to assess multiple technologies.

We were not able to track the time that each web-survey expert spent answering the assessment questions. We were, however, able to see whether an expert saved their answers during separate sessions, giving a partial indication of whether the experts split a survey into several sessions. We found only 2 experts used separate sessions spread over more than one day, suggesting that the time spend by web respondents was likely to be no greater than that spent by the interview participants. Finally, one of the interviewed experts commented that he felt the process used in the interview was a significant improvement over other surveys about similar issues to which he has been asked to respond. Specifically, he thought the discussions were useful, both to help him understand what we, as the analysts, were trying to accomplish and what kind of input we needed, and to both motivate him and allow him time to work through the assessments.

3. Discussion

3.1 Summary

In this paper we reviewed the literature on expert elicitations and combined this with a discussion of the differences between self-administered surveys and interviews for general public surveys. The literature on expert elicitation implies that interviews may be superior to self-administered, web-based surveys for a couple of reasons. When the focus of the elicitation is on developing estimates of specific uncertain quantities, a trained analyst can help experts to avoid some well-known systematic biases, such as overconfidence. More importantly, as the thinking about elicitations has evolved, a trained analyst may be crucial to help experts develop quantitative probability estimates based on their knowledge during the elicitation itself. This requires hard thinking and is greatly facilitated by the presence of an analyst. For elicitations with a broader focus, such as conceptual model design, the interactions between analyst and expert can be crucial to developing a shared understanding of the problem being modeled.

The literature on survey mode that has a clear relevance to EE is the prediction about cognitive effort. There is a hypothesis, and some evidence, that self-administered surveys require more cognitive effort and provide less motivation, leading to a likelihood of more satisficing behavior. In terms of EE, this may be likely to manifest itself as a higher level of systematic biases, as well as on probability distributions that are tossed off quickly rather than after the hard thinking we think it requires

We compared the results of two EE studies on CCS, one using interviews, the other web-based. The findings from this comparison are not conclusive, but are suggestive. First, we were surprised to find that among the most mature technologies (those that were judged to be viable by all experts under all scenarios), there was no evidence of a difference in the uncertainty range (or level of overconfidence) between the two modes. On the other hand, the overall amount of uncertainty expressed by the interview respondents was larger, as we expected, since many of these respondents reported that many of the technologies had a probability of infeasibility under some scenarios. This result is highly related to the process differences between the two modes, and in particular, to the flexibility provided by interviews. Some web-tools are being developed, however, that may increase the flexibility and interactivity of web-based modes,[29] thus reducing this gap.

We also found important differences between the survey modes in terms of the rankings of the technologies, and significant differences between the survey modes in terms of the assessed values. We are unable to determine whether or how much of these differences can be attributed to the elicitation mode and how much is due to the different samples – US versus EU.

This comparison gives weak support to the idea that web respondents are satisficing. The web respondents answered more questions and assessed more technologies than the interview respondents, while not appearing to take significantly more time to do so. In the survey literature this has been interpreted as evidence of satisficing, and we concur with this interpretation in this case. The interviews forced the respondents to think very hard about each of their answers, and to explain their thinking as they worked through the assessments. Their own time constraints and the cognitive load caused them to minimize the number of technologies assessed. The web-based EE required much more follow-up and clarification, which may be an indication that the experts moved quickly and less carefully through the questions. One of the web-based respondents included a discussion in the follow-up that suggested quite strongly that he had been “satisficing” in his responses – which were, according to the follow-up, to be interpreted as providing a general opinion about the effectiveness of the different policies, rather than as his actual estimates of what the EP would be for different technologies under different scenarios.

3.2 Future Research Agenda

Knowing if and when self-administered web-based elicitations, or new approaches using web-based facilitated elicitations, can be used as a substitute for in-person interview-based EE, would be extremely valuable, especially as the need for expert input for large society-wide problems such as government investments into energy technologies grows. This literature review and initial findings suggest a need for a future research agenda addressing this question.

First, while we believe that the ultimate question is how real experts, answering difficult questions about the future, perform under different modes, the level of ambiguity in the literature suggests that some controlled, lab-based experiments may be valuable. Perhaps previous, well-defined experiments on elicitation question design (e.g., Speirs-Bridge *et al.*[24]) could be extended and structured to compare modes in terms of quality of the responses, systematic biases, and the time spent pondering questions.

Second, it is crucial to follow up lab-based findings with studies using real EEs. A carefully designed study, avoiding sample bias, could provide results on whether there is a systematic difference between the types of values resulting from different modes. For example, we found that the web-based respondents were quite a bit more optimistic than the interviewees, but are not able to attribute that to the survey mode with any confidence. In addition, our finding that uncertainty ranges did not differ between the two modes is limited to the most mature, and least uncertain, technologies, and it would be valuable to see if, and under what conditions, that finding holds.

Thirdly, a study examining the amount of time spent on each question, as well as the amount of time researching questions off-line, could provide some insights into the level of satisficing.

As many of these large, society wide problems might have a global nature, web-based surveys could also provide a cost-effective and flexible tool to gather opinions from experts coming from a wider set of countries and cultures. Thus, it is relevant to assess whether cultural differences affect the performance of different survey modes in a systematic way.

References

1. National Reserch Council. Prospective Evaluation of Applied Energy Research and Development at DOE (Phase Two). Washington D.C.: The National Academies Press; 2007.
2. Baker E, Chon H, Keisler J. Advanced solar R&D: Combining economic analysis with expert elicitations to inform climate policy. Energy Economics. 2009a;31:37-49.
3. Baker E, Chon H, Keisler J. Carbon Capture and Storage: Combining Expert Elicitations to Inform Climate Policy. Climatic Change. 2009b;96(3):379-408.
4. 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):11497-11504.
5. Bosetti V, Catenacci M, Fiorese G, Verdolini E. The future prospect of PV and CSP solar technologies: An expert elicitation survey. Energy Policy. 2012; 49:308-317.
6. Fiorese G, Catenacci M, Verdolini E, Bosetti V. Advanced biofuels: Future perspectives from an expert elicitation survey. Energy Policy. 2013;56:293-311.
7. Spetzler CS, Stael Von Holstein C. Probability Encoding in Decision Analysis. Management Science. 1975;22:340-358.
8. Brier GW. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review. 1950: 78(1):1-3.
9. Winkler RL. The Assessment of Prior Distributions in Bayesian Analysis. Journal of the American Statistical Association. 1967;62(319):776-800.
10. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185:1124-1131.
11. Morgan MG, Henrion M. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York: Cambridge University Press; 1990.

12. Hora, SC. Assessing Probabilities from Experts in Edwards W, Miles RF, von Winterfeldt D (eds). Advances in Decision Analysis: From Foundations to Applications. Cambridge (UK): Cambridge University Press; 2007.

13. DeWispelare AR, Tandy Herren L, Clemen RT. The use of probability elicitation in the high-level nuclear waste regulation program. International Journal of Forecasting. 1995;11:5-24.

14. Fischhoff B. Eliciting Knowledge for Analytical Representation. IEEE Transactions on Systems, Man, and Cybernetics. 1989;19(3):448-461.

15. Cooke RM. Experts in uncertainty: Opinion and subjective probability in science. New York: Oxford University Press; 1991.

16. SSHAC (Senior Seismic Hazard Analysis Committee). Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts. Washington DC (US): Nuclear Regulatory Commission; 1997. Report No.: NUREG/CR-6732.

17. Linstone HA, Turoff M. The Delphi method: Techniques and applications. Boston (MA): Addison-Wesley; 1975.

18. Jenni K, van Luik A. Assessment of expert judgments. In: Ahn J, Apted M (eds). Geological Repository Systems for Safe Disposal of Spent Nuclear Fuels and Radioactive Waste. Boca Raton (FL): CRC Press; 2010: 580–609.

19. Kotra JP, Lee MP, Eisenberg NA, DeWispelare AR. Branch Technical Position on the Use of Expert Elicitation in the High-Level Radioactive Waste Program. Washington DC (US): Nuclear Regulatory Commission; 1996. Report No.:NUREG-1563.

20. Cooke RM, Goossens LHJ. Procedures guide for structured expert judgment. Brussels-Luxembourg: Commission of the European Communities; 1999. Report No.:EUR18820

21. Meyer MA, Booker JM. Eliciting and Analyzing Expert Judgment: A Practical Guide. Philadelphia (PA): Society for Industrial and Applied Mathemathics (SIAM); 2001.

22. Nordhaus WD. Expert opinion on climatic change. American Scientist. 1994;82:45–51.

23. Curtright A, Morgan MG, Keith D. Expert assessment of future photovoltaic technology. Environmental Science and Technology. 2008;42:9031-9038.

24. Spiers-Bridge A, Fidler F, McBride M, Flander L, Cumming G, Burgman M. Reducing overconfidence in the interval judgments of experts. Risk Analysis. 2010a;30(3):512-523.

25. 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;4:2685-2692.

26. James A, Choy SL, Mengersen K. Elicitator: An expert elicitation tool for regression in ecology. Environmental Modelling & Software. 2010;25(1):29–145.

27. Spiers-Bridge A, Flander L, Robinson A. Expert elicitation using real-time expert panels. A presentation at the annual conference of the Society for Risk Analysis, 2010b. Available at: http://www.acera.unimelb.edu.au/sra/2010/Presentations/Speirs-Bridge.pdf

28. Spaccasassi C, Deleris L. A Web-based Tool for Expert Elicitation in Distributed Teams. Proceedings of the 8th Bayesian Modelling Applications Workshop of UAI 2011, 2011. Available at: http://abnms.org/uai2011-apps-workshop/papers/SpaccasassiDeleris.pdf

29. Inman M. Near Zero, How Low Will Photovoltaic Prices Go? An Expert Discussion. 2012 Dec. Available at: http://www.nearzero.org/reports/pv-learning.

30. Dalal S, Khodyakov D, Srinivasan R, Straus S, Adams J. ExpertLens: A system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge. Technological Forecasting and Social Change. 2011;78(8): 1426–1444.

31. Bowling A. Mode of Questionnaire Administration Can Have Serious Effects on Data Quality. Journal of Public Health. 2005; 27:281–91.

32. Simon HA. Rational choice and the structure of the environment. Psychological Review. 1956;63(2):129-138.

33. Krosnick JA. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. Applied Cognitive Psychology. 1991;5:213–36.

34. Fricker S, Galesic M, Tourangeau R, Yan T. An experimental comparison of web and telephone surveys. Public Opinion Quarterly. 2005;69(3):370-392.

35. Heerwegh D, Loosveldt G. Face-to-Face versus Web Surveying in a High-Internet-Coverage Population. Differences in Response Quality. Public Opin Q. 2008;72(5):836-846.

36. Marta-Pedroso C, Freitas H, Domingos T. Testing for the survey mode effect on contingent valuation data quality: a case study of web based versus in-person interviews. Ecological Economics. 2007;62:388-398.

37. Leggett CG. Social desirability bias in contingent valuation surveys administered through in-person interviews. Land Econ. 2003;79:561–575.

38. Nielsen JS. Use of the Internet for Willingness-to-pay Surveys. A Comparison of Face-To-Face and Web-Based Interviews. Resource and Energy Economics. 2011;33(1):119-129.

39. Taylor PA, Nelson NM, Grandjean BD, Anatchkova B, Aadland D. Mode effects and other potential biases in panel-based Internet surveys: Final report. Laramie (WY): Wyoming Survey & Analysis Center, University of Wyoming; 2009. Report No.: SRC-905.

40. Maguire KB. Does mode matter? A comparison of telephone, mail, and in-person treatments in contingent valuation surveys. J Environ Management. 2009;90(11):3528-33.

41. Covey J, Robinson A, Jones-Lee M, Loomes G. Responsibility, scale and the valuation of rail safety. Journal of Risk and Uncertainty. 2010; 40(1):85-108.

42. Lindhjem H, Navrud S. Are Internet Surveys an Alternative to Face-to Face Interviews in Contingent Valuation?. Ecological Economics. 2011;70(9):1628-1637.

43. Dillman D, Phelps G, Tortora R, Swift K, Kohrell J, Berck J, Messer B. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the internet. Soc. Sci. Res. 2009;38:1–18.

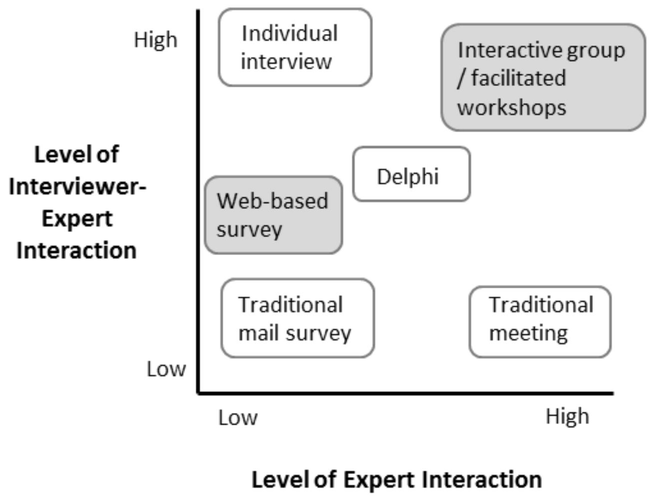
44. Jenni K, Baker E, Nemet GF. Expert elicitations of energy penalties for carbon capture technologies. International Journal of Greenhouse Gas Control. 2013;13:136-145.

45. Nemet G, Baker E, Jenni K. Modeling the future costs of carbon capture using experts' elicited probabilities under policy scenarios. Energy. 2013**;** 56(1): 218-228.

46. Clemen RT, Winkler RL. Combining Probability Distributions from Experts in Risk Analysis. Risk Analysis. 1999;19(2):187-200.

47. Cardenas IC, Al-jobouri SSH, Halman JIM, van Tol FA. Capturing and Integrating Knowledge for Managing Risks in Tunnel Works. Risk Analysis. 2013:33(1):92-108.

Figures and Tables



(This figure is based on Chapter 7, Figure 2, p. 102 of Meyer and Booker, 2001. Grey shaded node represent new or modified elements of the original figure)

Figure 1. Level of interaction in different elicitation modes

Please fill in the table below with your assessments of the low, central, and high estimates **for the energy penalty associated with carbon capture using Absorption/Solvents technologies** in 2025, under the three different scenarios.

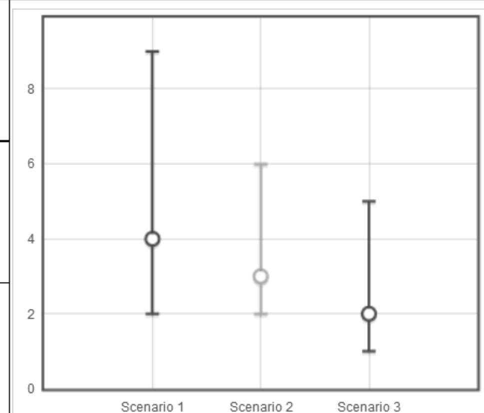
	Scenario 1	Scenario 2	Scenario 3	LIVE PLOT
High estimate: 1/20 chance that energy penalty will be higher than this	9	6	5	
Median	4	3	2	
Low estimate: 1/20 chance that energy penalty will be lower than this	2	2	1	
Please provide here the rationale for your answers	comments	comments	comments	

Figure 2: Snapshot of the questions in the web survey

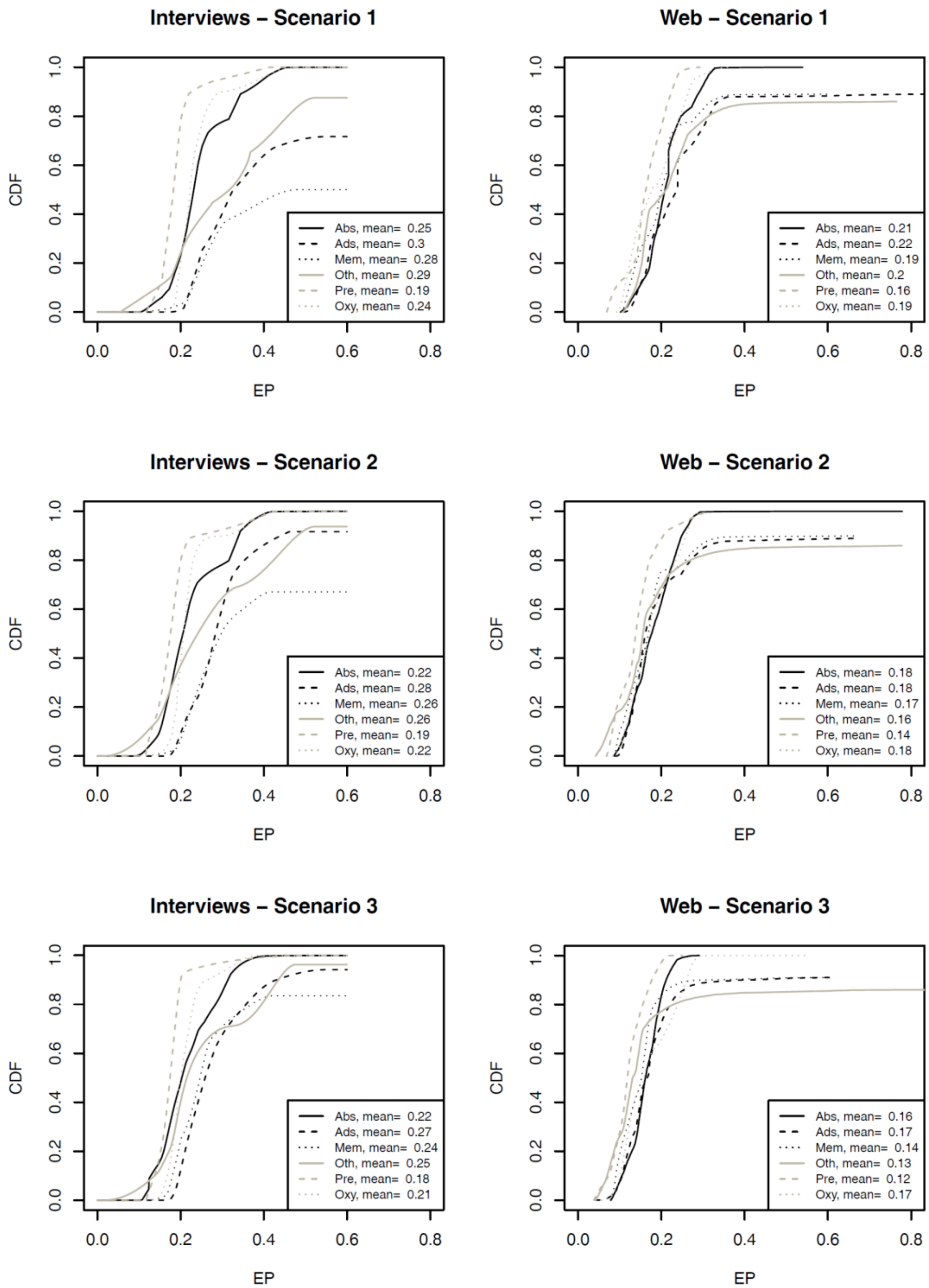


Figure 3: Aggregated cumulative distribution functions and mean value for the energy penalty of each carbon capture technology resulting from the expert interviews (left panels) and web-based surveys (right panels) surveys, for each scenario.

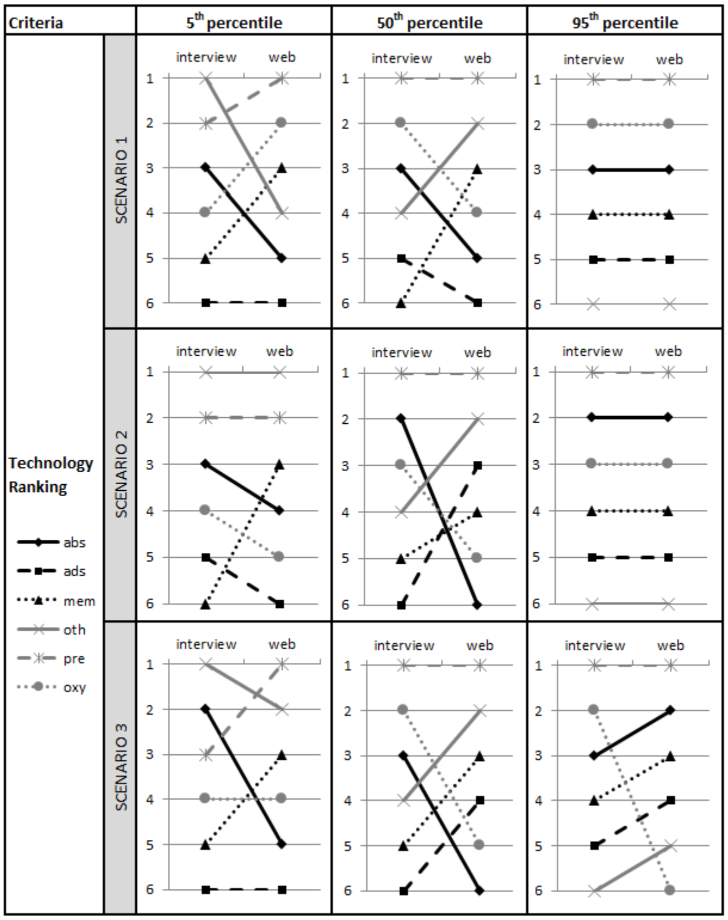


Figure 4: Comparison of technology rankings by elicitation mode, for three scenarios and three ranking indices.

Table I. Steps in an expert elicitation study.

Steps in conducting an expert elicitation study
1. Define the objectives of the study and determine whether expert elicitation is necessary and appropriate for meeting the study needs. Determine how the elicitation is to be carried out.
2. Select the experts.
3. Structure the assessment, identify and clarify the assessment issues, develop any assessment protocols necessary for the elicitation method. In some approaches this step is carried out by the study team separate from any interactions with the experts who will be the elicitation participants, in others it is included as part of the elicitation process.
4. Develop and provide the experts with background and training about assessment tasks, cognitive biases, and probability encoding concepts.
5. Conduct the elicitation itself and provide feedback about their elicitation results and their implications. In some approaches, this includes the opportunity for the experts to revise or update their assessment
6. Use the individual expert's inputs to create an aggregate assessment of the quantities of interest, if desired. The results of the aggregation can be included in further feedback to the experts.
7. Document the assessment and results

Table II. Results of studies comparing survey modes.

	Evidence for "satisficing"		
Study	Lower response variability	Higher number of non- or DK-responses[1]	Lower time to complete
Fricker et al. (2005)	Self	Interview	Interview
Heerwegh and Loosveldt (2008)	Self	Self	Self
Lindhjen and Havrud (2011)	No difference	No difference	Not tested

[1] Differences in the number of non-responses in the first two studies may be attributable to the ease of accessing a DK response (e.g., visible in the self-administered survey but not offered the interview; confirmation of DK required in the self-administered survey)

Table III: OLS regression factors.

FACTOR	LEVEL	DEFINITION
Technology	1	Absorption*
	2	Oxyfuel
	3	Pre-combustion
Policy scenario	1	No further government R&D*
	2	Carbon policy / carbon pricing
	3	Increased R&D investments
Survey	1	Web-based*
	2	Interview

* indicates the reference case

Table IV: Results of different OLS models for coefficient of variation of fitted distributions

	Coefficient of Variation	
	Model 1	**Model 2**
Intercept	0.11669 ***	0.11061 ***
Technology effect	no	no
Scenario 2 (versus S1)	0.01656	0.01656 .
Scenario 3 (versus S1)	0.03141 *	0.03141 ***
Survey (interviews vs web-based)	-0.01435	0.00741
Expert effect	-	yes

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table V: Results of OLS regression for model of equation (2) on expert answers.

	5%	**50%**	**95%**
Intercept	0.17607 ***	0.20075 ***	0.23734 ***
Technology effect	yes	yes	yes
Scenario 2 (versus S1)	-0.02191 *	-0.02037 *	-0.01876 .
Scenario 3 (versus S1)	-0.03486 ***	-0.03611 ***	-0.02942 **
Survey (interviews vs web-based)	0.03589 ***	0.04425 ***	0.05612 ***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table VI: Results of OLS regression for models of equation (3) on expert answers.

	5%	**50%**	**95%**
Intercept	0.17419 ***	0.19969 ***	0.23720 ***
Technology effect	yes	yes	yes
Scenario 2 (versus S1)	-0.02191 *	-0.02037 *	-0.01876 .
Scenario 3 (versus S1)	-0.03486 ***	-0.03611 ***	-0.02942 **
Survey (interviews vs web-based)	0.01552	0.01465	0.02894 .
Country: US vs EU	0.02577 .	0.03625 *	0.03270 .

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Appendix

To assess the robustness of the findings related to the OLS regressions on the expert answers on the percentiles of their distribution of the future EP, we performed some additional regression analyses. We add a random effect to control for the individual experts, so that the resulting model is a mixed effect model:

$$EP=\alpha+\beta_{tech}\cdot technology+\beta_{scen}\cdot scenario+\beta_{surv}\cdot survey+\delta_{expert}+\varepsilon\,, \tag{4}$$

Where δ_{expert} is an expert specific error term, that captures the expert variability.

Table VII reports the results of this regression. Most of the results of the models included in the main text of the paper are confirmed in size and level of significance. What changes is the significance level at which the survey effect is statistically different from zero. Though it is important that the effect has remained statistically significant and with similar values. The lower level of significance may be explained by the fact that in the previous model we were treating different answers from the same expert as independent, while now we are putting them in relation, reducing the effective sample size.

	5%		50%		95%	
Intercept	0.17678	***	0.20235	***	0.23690	***
Technology effect	yes		yes		yes	
Scenario 2 (versus S1)	-0.02191	*	-0.02037	*	-0.01876	.
Scenario 3 (versus S1)	-0.03486	***	-0.03611	***	-0.02942	
Survey (interviews vs web-based)	0.03129	*	0.04358	*	0.06328	*
Expert effect	yes		yes		yes	

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table VII: Results of OLS regression for model of equation (4) on expert answers, percentiles

We also test a model that includes an interaction effect ($\gamma_{scen,surv}$) between the scenario factor and the survey dummy:

$$EP=\alpha+\beta_{tech}\cdot technology+\beta_{scen}\cdot scenario+\beta_{surv}\cdot survey+\gamma_{scen,surv}\cdot scenario\cdot survey+\varepsilon\,. \tag{5}$$

The resulting model allows us to assess whether the scenario effect is different between our two samples of respondents (survey mode). The interaction effect is not significant, as shown in the results reported in Table VIII.

	5%		50%		95%	
Intercept	0.18256	***	0.20554	***	0.24129	***
Technology effect	yes		yes		yes	
Scenario 2 (versus S1)	-0.02821	*	-0.02430	.	-0.02130	
Scenario 3 (versus S1)	-0.04802	***	-0.04655	**	-0.03873	*
Survey (interviews vs web-based)	0.02400	.	0.03546	*	0.04888	**
Interaction: S2·interview	0.01155		0.00720		0.00467	
Interaction: S3·interview	0.02413		0.01915		0.01707	

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table VIII: Results of OLS regression for model of equation (5) on expert answers, percentiles

The two samples of experts appear to agree on the impact of the policy with respect to scenario 1, i.e., the difference between S1 and S2, and S1 and S3, and S2 and S3 is not dependent on the survey type. Finally, we perform one last robustness check by applying all models (reported in equations 2, 3, 4,and 5) first to the mean value of the fitted distributions, rather than the percentile answers of the experts. Table IX reports the results for all the models, reported on the various columns, which confirm our results.

	Mean of fitted probability distribution							
	Eq 2 model		**Eq 3 model**		**Eq 4 model**		**Eq 5 model**	
Intercept	0.21595	***	0.21584	***	0.21846	***	0.22179	***
Technology effect	yes		yes		yes		yes	
Scenario 2 (versus S1)	-0.02089	*	-0.02089	*	-0.02089	***	-0.02688	*
Scenario 3 (versus S1)	-0.03501	***	-0.03501	***	-0.03501	***	-0.04653	***
Survey (interviews vs web-based)	0.03686	***	0.01100		0.03433	.	0.02595	*
Country: US vs EU			0.03108	*				
Country: Other vs EU			0.00197					
Experct effect					yes			
Interaction: S2·interview							0.01120	
Interaction: S3·interview							0.02152	

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table IX: Results of different OLS models for mean of fitted distributions

NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI

Fondazione Eni Enrico Mattei Working Paper Series

Our Note di Lavoro are available on the Internet at the following addresses:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=266659
http://ideas.repec.org/s/fem/femwpa.html
http://www.econis.eu/LNG=EN/FAM?PPN=505954494
http://ageconsearch.umn.edu/handle/35978
http://www.bepress.com/feem/

NOTE DI LAVORO PUBLISHED IN 2014

CCSD	1.2014	Erin Baker, Valentina Bosetti, Karen E. Jenni and Elena Claire Ricci: Facing the Experts: Survey Mode and Expert Elicitation