

Energy Technology Expert Elicitations for Policy: Their Use in Models and What Can We Learn from Workshops and Meta-analysis

Laura Diaz Anadon (Harvard University),* Valentina Bosetti (Bocconi University and Fondazione Eni Enrico Mattei), Gabriel Chan (Harvard University), Gregory Nemet (University of Madison-Wisconsin), Elena Verdolini (Fondazione Eni Enrico Mattei)

Working Paper prepared for:

Methods for Research Synthesis: A Cross-Disciplinary Workshop

Harvard Center for Risk Analysis

October 3, 2013

www.hcra.harvard.edu

*Corresponding author: laura_diaz_anadon@harvard.edu

Disclaimer: The findings and conclusions of this paper are those of the authors and do not imply endorsement by any of their affiliations or funders. Comments should be directed to the authors.

Acknowledgements: Laura Diaz Anadon and Gabriel Chan acknowledge funding from the Climate Change Initiative of the Doris Duke Charitable Foundation, and a grant from BP-International Limited on Energy, Climate & Security Policy. Valentina Bosetti and Elena Verdolini received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 240895 – project ICARUS “Innovation for Climate Change Mitigation: a Study of energy R&D, its Uncertain Effectiveness and Spillovers.” Gregory Nemet acknowledges support from the Wisconsin Alumni Research Foundation.

**Energy technology expert elicitations for policy:
their use in models and what can we learn from workshops and meta-analysis**

Laura Diaz Anadon,^{a*} Valentina Bosetti,^{b,c} Gabriel Chan,^a Gregory Nemet,^d and Elena Verdolini^c

^aBelfer Center for Science and International Affairs, Harvard Kennedy School, Harvard University

^bDepartment of Economics, Bocconi University, Milan, Italy

^cFondazione Eni Enrico Mattei, Milan, Italy and CMCC, Lecce, Italy

^dLaFollette School of Public Affairs, University of Madison-Wisconsin

*Corresponding author

Abstract

Characterizing the future performance of energy technologies can improve the development of energy policies that have net benefits under a broad set of future conditions. In particular, decisions about public investments in research, development, and demonstration (RD&D) that promote technological change can benefit from (1) an explicit consideration of the uncertainty inherent in the innovation process and (2) a systematic evaluation of the tradeoffs in investment allocations across different technologies. To shed light on these questions, over the past five years several groups in the United States and Europe have engaged in research initiatives, developing insights on technology forecasts based on expert elicitations and integrating these results in energy-economic models. In this paper, we discuss the lessons learned from the design and implementation of these initiatives in four respects. First, we discuss lessons from the development of ten energy-technology expert elicitations protocols, highlighting the need for (and difficulties associated with) matching elicitation design with a particular modeling tool. Second, we present insights from the use expert elicitations to optimize RD&D investment portfolios. These include a discussion of decreasing marginal returns to research, of the optimal level of overall investments, and of the sensitivity of results to policy scenarios and to the selected metrics for evaluation. Third, we discuss the effect of combining online elicitation tools with group discussions on ability of researchers to utilize the results. Fourth, we summarize the results of a meta-analysis of elicited data across research initiatives to identify the impact of expert selection on elicitation results and the associated expected returns to RD&D.

Keywords

Expert elicitations, energy technology innovation, public R&D, meta-analysis, energy modeling, optimization

1. Introduction

Public investments in energy technology research, development and demonstration (RD&D) have been justified by governments throughout the world on the basis of public policy challenges that fall in three broad categories (environmental externalities, energy security, and economic competitiveness) (Anadon, 2012) in addition to the knowledge spillovers associated with scientific research more generally (Arrow, 1962). Country members of the International Energy Agency¹ invested \$13.7 billion PPP in public energy RD&D in 2008, which rose to \$17 billion PPP (Purchasing Power Parity) in 2012 (IEA, 2013). In 2012, the United States alone invested just over \$4.7 billion PPP, while European countries' invested totaled \$5.8 billion PPP. Data for emerging economies is scarce, but a recent review of the largest developing countries (Brazil, Russia, India, Mexico, China, and South Africa) indicates that in 2008, public energy RD&D was of a comparable scale to IEA countries, totaling \$13.8 billion PPP (Gallagher, Anadon, Kempener, & Wilson, 2011).

Energy RD&D investments are a fraction of energy deployment subsidies in dollar terms.² However, the impact of these RD&D investments may be proportionally larger than that of subsidies because of the long-term, fat-tailed, and generally high nature of the benefits associated with the innovation process (Nemet, 2013). Based on this view, since 1996 many expert panels in the United States (American Energy Innovation Council, 2010; NCEP, 2004; NCEP, 2007; PCAST, 1997; PCAST, 2010) and in the European Union (EERA, 2010; European Commission,

¹ The IEA has 28 Member countries (Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Republic of Korea, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States), but no data are provided for Luxembourg or the Slovak Republic. Iceland, Chile, and Mexico are OECD members, but are not IEA members.

² The United States government spent about \$33.2 billion in 2010 in energy subsidies for deployment (IEA, 2011). A recent report estimated that total energy subsidies for deployment in 2007 were \$483 billion (IEA, OPEC, OECD, & World Bank, 2010).

2007) have called for significant increases to public investments in energy RD&D. These studies, however, offer little analytic support to justify their recommendations and often do not include careful estimates of benefits and costs.

The U.S. Department of Energy (DOE), the single funder of energy RD&D in the United States, often conducts estimates of the expected benefits of individual RD&D programs. However, the DOE does not consistently evaluate the interactions of its programs across its investment portfolio (e.g., the fact that storage may complement intermittent renewables, and that nuclear and carbon capture and storage could act as substitutes), which may be positive or negative, nor does it systematically consider uncertainty in its benefit calculations. In those benefit assessments that the DOE does conduct, a lack of transparency in the source of its inputs (usually internal) leads to questioning the credibility of its estimates. In short, the DOE does not conduct robust cost-benefit analysis to support its portfolio of RD&D investment decisions in different technology programs in a robust, consistent, or transparent way. As a result of some of these shortcomings, a 2007 study of the National Research Council recommended that the DOE make probabilistic assessments of the impacts of RD&D programs when making decisions (NRC, 2007).³

The political economy conditions within an RD&D funding organization make generating credible estimates of the impact of RD&D more difficult. For example, in the case of DOE, competition between the different technology programs creates incentives for self-serving biases

³ Different strands of research have tried to estimate some of the returns to society of supporting energy RD&D using more analytical approaches. Schock et al. (1999) (Schock et al., 1999) and Nemet and Kammen (2007) (Nemet & Kammen, 2007) estimated the appropriate level of energy R&D as the difference between the cost of meeting CO₂ emissions targets using assumptions of business-as-usual and advanced technology costs. Davis and Owens (2003) (Davis & Owens, 2003) estimate the value of investments in renewable energy R&D using real options. These three papers do not provide insights regarding the allocation of RD&D funds across different energy technologies. The methodology presented by Blandford (2009) (Blandford, 2009) addresses the allocation question by estimating the optimal allocation of R&D funds for renewable energy, nuclear energy, and coal with CCS by defining two states for the cost of those technologies (BAU and low), assuming that the probability of achieving the low-cost technology is an exponential function of R&D. These studies however do not rely on transparent, consistent, and credible (unbiased) assumptions about the expected impact of public RD&D on the future cost and performance of energy technologies. Another approach has been to leverage historical data to forecast future technical change (McNerney, Farmer, & Trancik, 2011). However, this approach cannot capture the fact that future innovation may proceed through unprecedented pathways, making the past a poor predictor of the future. This means that, even though the insights from these studies have been valuable, their results cannot be directly used in actual policy decisions about optimal levels of energy RD&D investments and optimal allocation.

and erodes trust between programs. . One strategy that appears to meet these political-economy conditions is eliciting the knowledge from external experts and integrating this knowledge into internally-acceptable assessment frameworks (Chan & Anadon, 2013).

In this vein, two groups—one at the Harvard Kennedy School, HKS, and one at the Fondazione Eni Enrico Mattei, FEEM—have over the past five years have utilized expert elicitations in studies that estimate the relationship between public RD&D investments and technology cost and performance both in the United States and the European Union (Anadon et al., 2011a; Anadon et al., 2011a; Anadon, Bosetti, Bunn, Catenacci, & Lee, 2012; Anadon et al., 2012; Bosetti, Catenacci, Fiorese, & Verdolini, 2012; Bosetti et al., 2012; Catenacci M., Verdolini E., Bosetti B., & Fiorese G., 2013; Chan, Anadon, Chan, & Lee, 2011; Fiorese, Catenacci, Verdolini, & Bosetti, 2013). Some similar studies also utilized energy technology expert elicitations, but were not explicitly developed to provide insights about portfolios of investments at a large scale (e.g., for technology programs funded by DOE or the EU Commission) or across multiple technologies (Baker, Chon, & Keisler, 2009a; Baker, Chon, & Keisler, 2008; Baker, Chon, & Keisler, 2009b; Curtright, Morgan, & Keith, 2008).

In light of the wealth of research conducted in the past few years in this respect, this paper discusses lessons learned regarding how expert elicitations can be designed, implemented, and utilized to support decisions about the allocation of public energy RD&D investments from results presented in (Anadon et al., 2011b) (Anadon et al., 2011a; Anadon et al., 2012; Anadon, Nemet, & Verdolini, 2013), and supported by the additional expert elicitation papers of the HKS and FEEM groups cited above. These studies incorporate findings from a large set of expert elicitations designed with the objective of providing insights to both the DOE and EU policy makers about the allocation of funding for nuclear power, solar photovoltaics and concentrated solar power, biofuels and bioelectricity, utility scale energy storage, fossil power with and without carbon capture and storage, and vehicles. Elicitations for the US were carried out by researchers at HKS between 2009 and 2011 and were designed to allow the use of their results in MARKAL (Fishbone & Abilock, 1981), a widely used energy-economic model, to provide insights about DOE funding decisions across programs. Elicitations for the EU were carried out by Fondazione Eni Enrico Mattei within the FP7 project ICARUS and designed for use in WITCH (www.witchmodel.org), and integrated assessment energy model. Finally, this paper

also includes insights and findings from a meta-analysis of the nuclear technology elicited data conducted by researchers at Harvard, FEEM, and the University of Madison-Wisconsin aimed at identifying how elicitation design affects estimates of the impact of R&D on future technical change.

The rest of the paper is structured as follows. Section 2 presents a literature review on the previous use of expert elicitations for energy technologies. Section 3 describes the design and implementation of expert elicitations in an energy-economic modeling context (MARKAL and WITCH) conducted by the Harvard and the FEEM groups, respectively. Section 4 discusses the insights from conducting elicitations online and complementing the elicitations with a group workshop. Section 5 discusses the insights from utilizing expert elicitations as modeling inputs to estimate the outcomes of RD&D investments. Section 6 discusses the insights from combining results from various elicitations in one technology area (nuclear power). Section 7 concludes with a summary of findings and thoughts for future research.

2. Expert Elicitations of Energy Technologies to and RD&D investment decisions

Estimating the impact of energy RD&D investments requires estimation of two relationships. First is the relationship between particular RD&D investments and individual technology outcomes, which are typically measured in terms of cost but also of performance. Second is the relationship between the technology outcomes and policy goals, such as economic growth, energy prices, CO₂ emissions, or oil imports.

Expert elicitations are being increasingly used to estimate the first relationship. These studies gather the opinions of experts on technical questions that fall within their area of knowledge and expertise. Data collection is carried out using elicitation protocols carefully designed to reduce heuristics and biases (Cooke, 1991; Hogarth, 1987; Morgan & Henrion, 1990). These data-gathering efforts are particularly useful in decisions that require an assessment of the future evolution of energy technologies because historic data may not inform future performance and costs or because the relevant data might not be available. However, few studies have designed elicitations with the objective of supporting specific energy RD&D policy decisions on a

continuous basis. In addition, even though previous studies have indicated the importance of protocol design and expert selection as key for elicitation results, (Keeney & Winterfeldt, 1991; Meyer & Booker, 1991; Raiffa, 1968) there are no empirical assessments of the impact and size of differences in elicited results from expert selection and elicitation design (e.g., whether the survey is conducted in person, via mail, or online).

Results from expert elicitations can also be used to estimate the second relationship through an integrated assessment that links elicited relationships between the first relationship (between RD&D investments and technology outcomes) in a framework that models the second relationship (between technology outcomes and policy goals). By propagating the uncertainty from the first relationship through the second relationship, such an analysis can provide important insights on expected benefits of RD&D investments and their uncertainties. Specifically, expert elicitation estimates of future technology cost and performance can be introduced, if properly designed, in technologically-detailed models of the economy to link technology outcomes to social benefits, improving the ability of decision makers to understand how technological uncertainty propagates through to benefits in the market place.

3. Methods

3.1 Design and implementation of expert elicitations

In this paper we present findings from a Harvard study designed to inform the U.S. Department of Energy decisions on the allocation of RD&D investments across large scale technology programs (nuclear, solar, vehicles, etc.), and from part of a European Research Council funded project (ICARUS) that aimed at designing optimal allocation of the EC research budget on energy technologies, with a specific attention to the role of European climate and energy policies. With this goal in mind, below are some of the key features of both the Harvard and FEEM elicitations, which were conducted between 2009 and 2011.

Both institutions conducted six elicitations each. The Harvard study conducted six elicitations in two media, with four of the elicitations on paper and distributed by mail (bioenergy, utility scale storage, fossil energy and carbon capture and storage, and vehicles) and two elicitations online (nuclear power and solar PV). The European project conducted six expert elicitations, of which

four were conducted through extensive in person interviews (batteries for EDV, bioenergy, biofuels, solar) and two were conducted online (carbon capture and storage, and nuclear power).

For both research groups, the core objective of the elicitations was to gain insights about the relationships between public RD&D investments and technological change in specific technologies in a parameterization that could be naturally introduced into an economic model of aggregate benefits. Specifically, the Harvard elicitations generally included questions about the experts' estimates about various cost components (e.g., overnight capital cost, operations and maintenance costs, costs of important components) as well as different performance parameters (e.g., efficiency, yield, fuel efficiency), both of which were tailored to the different technologies, in 2030 under different DOE RD&D budgets. The two exceptions were the bioenergy survey—in which experts were given the option of providing a cost breakdown or providing an overall cost per unit of biofuel or electricity delivered—and the vehicles survey—in which experts were asked about the total purchasing cost of different types of vehicles and specific performance characteristics without a breakdown of cost components (e.g., battery cost). The FEEM elicitations for batteries for electric vehicles, bioelectricity, biofuels, and solar power asked experts to provide an aggregated metric of the 2030 cost of a particular technology under different E.U. RD&D budgets. The FEEM carbon capture and storage (CCS) survey investigated both the cost and energy penalty of alternative CCS technologies. Finally, the FEEM nuclear survey was conducted in coordination with the Harvard study and using the two step methodology including an online individual elicitation of U.S. and E.U. experts followed by a workshop with a subset of experts participating in the individual nuclear surveys and also asked questions about cost components and different performance parameters. (For more detailed information the readers are referred to the papers on the Harvard and FEEM elicitations provided in section 1).

The choice of the media for conducting the elicitation is an important one and generally thought of as involving tradeoffs. Previous elicitation studies have stressed the possible benefits of in-person interviews, as they allow the interviewer and the expert to interact, thereby allowing researchers to pose follow up questions and remind experts throughout to try to reduce their bias and use of availability heuristics, among other possible benefits. However, conducting the elicitations through the mail and online reduces the monetary and time cost of the elicitation process and increases flexibility for both the research team and the participating experts; thereby

increasing the feasibility of conducting elicitations for to support policy decisions and expanding the pool of participants.

Independent of the media chosen to administer the elicitation, the process of developing the elicitation protocol and instrument for the first time for both research groups was around 3-5 months and was consistent with previous energy technology expert elicitations. This development process included the crucial step of testing and revising the elicitation protocol through pilot interviews and revise each elicitation device in an iterative process with two to three experts in that technology.⁴

The elicitations followed the protocol of previous elicitations in the literature by including a background calibration section before presenting the main elicitation questions. The background section contained a summary of the purpose of the survey, background information on either DOE's or EU current activities and investments in the technology area of interest, and a statement about avoiding bias and overconfidence.⁵ All Harvard surveys and most FEEM's surveys also asked participants to rate their own expertise in several sub-technology areas on a 6-point scale, where 6 was described as "I am one of the top experts in this technology/system" and 1 was described as "I am not familiar with this technology/system." This information was subsequently used to test for correlations between areas of expertise and either differential recommendations for RD&D funding or particularly optimistic technology forecasts, which would have been consistent with experts making self-interested recommendations⁶.

The second half of the protocol contained the core questions of the elicitations. Regarding the Harvard studies, this part included four sections: (1) questions about the commercial viability, cost and performance of different technologies in 2030 under a business as usual (BAU) public RD&D funding scenario; (2) questions about the expert's recommendation of total public investments in the general technology area of the survey and their recommended allocation of funds to sub-technologies, including questions about the specific technical hurdles to be addressed by their allocation; (3) questions about how future technology costs and performance

⁴ It is worth noting that even though experts filled out the elicitations on their own with a few exceptions (one filled out one survey with one of the researchers and two filled them out with one of the researchers on the phone), experts were encouraged to contact researchers throughout the process with any questions that arose.

⁵ The 100 elicitation participants that participated in the Harvard surveys were identified through peer-reviewed publications, National Academies reports, participation in conferences, and referrals from other experts. Experts received between one and three invitations to participate.

⁶ We also considered using self-rated expertise to weight experts, but we ultimately did not conduct this analysis.

would change if their recommended RD&D investments were implemented, and how this would change under alternative RD&D investment levels⁷; and (4) qualitative technology-specific questions about other policies and factors affecting technology deployment, which differed significantly across the 6 elicitations. With respect to the FEEM elicitations on batteries for EDV, bioenergy, biofuels, and solar experts were asked to (1) assess different technological options based on their level of maturity and possible bottleneck; (2) suggest a breakdown of public research expenditures across the different technological paths within each different elicitation that would maximize the change of a breakthrough in that technology; (3) provide estimates of future costs and the surrounding uncertainty conditional on different levels of public RD&D investment;⁸ (4) assess the potential bottlenecks associated with a given technology that were additional RD&D investment could not address (such as, for example, concerns about the competition of biofuel with food for given crops) and (5) assess the potential international diffusion (to both OECD and non-OECD countries) of a given technology after reaching a breakthrough in costs.

All elicitations included interactive visual aids. The Harvard mail surveys included a set of chips and a “board game” to help experts think through allocating their recommended budget across different technology areas and technology development “stages”.⁹ The Harvard and FEEM online surveys included a virtual game board and chips as well as graphical feedback for all quantitative input from the experts, allowing them to visualize their probability distribution of their cost and performance estimates under the different RD&D scenarios under investigation.¹⁰ The FEEM in person surveys allowed experts to plot their cost estimates in real time and check for the consistency of their own answer.

⁷ We asked experts to reassess future technology cost and performance metrics if their full recommended investment strategy were to be implemented, as well as strategies with the same proportional allocation but with half of their recommended aggregate budget and with ten-times their recommended aggregate budget.

⁸ The FEEM elicitations asked the expert to first provide estimates of the 10th, 90th and 50th percentile of future costs. The same experts were subsequently asked to provide probabilities that under the same different RD&D scenario the cost of a given technology would be below some level chosen by the researchers. This effectively meant eliciting the same information twice, but under different format, and allowed to check the consistency of expert’s responses.

⁹ The Harvard board game included 100 poker chips, one for each percentage of their total recommendation, that experts allocated across sub-technology areas, which included an “other category” that allowed them to indicate additional areas. The stages of RD&D that experts could allocate across were basic research, applied research, pilots, and demonstration.

¹⁰ The graphical feedback on the online surveys included plots of the 90th, 10th, and 50th percentile estimates for each technology and different budget scenarios, allowing experts to modify their answers as they were filling out the graphs in real-time.

To evaluate the extent to which conducting elicitations online was effective, the FEEM and Harvard groups partnered to conduct nearly identical elicitations in nuclear energy. Following the surveys, the groups convened a subset of the European and U.S. experts for a 1.5-day workshop to discuss the results of the survey and to bring forward any questions or misunderstandings that may have surfaced when completing the elicitations. Experts were asked to discuss their answers and to talk through their disagreements regarding the interpretation of the questions. Following each session of the workshop, experts were given the opportunity to privately change their answers.

Finally, both research groups worked at connecting the technical outcomes and/or the cost and uncertainty estimates obtained in the elicitations to societal benefits (e.g., CO₂ emissions, energy costs, oil imports, etc.). The MARKAL model was selected by the Harvard group, while the FEEM group worked with the WITCH model. MARKAL is a bottom up energy-economic model that is publicly-available and has institutional buy-in from many government agencies in the United States and elsewhere. The Harvard team coupled the use of the model with an importance sampling technique. This importance sampling methodology allowed to adjust for changing input assumptions without requiring additional model runs, thus solving a computational constraint faced by many decision-making entities¹¹ (Pugh et al., 2011). Because of our method's ability to test different input assumptions, the benefits associated with RD&D investments under the assumptions of different experts, some more pessimistic and some more optimistic, were estimated. The method can be used to conduct other sensitivity analysis such as including experts internal to the decision making process vs. experts from stakeholder groups, experts from different countries, etc (Chan & Anadon, 2013). It also can be used to understand the sensitivity of aggregated results to decisions about whether to include or exclude the outlier expert responses (Jenni, Baker, & Nemet, 2013).

¹¹ The computational challenge comes after the challenge of building internal trust and buy-in, achieving external transparency and consistency, which currently contributes to decision-making entities not estimating the benefits of RD&D investment portfolios.

3.2 Meta-analysis of expert elicitations

Given the scarcity of information regarding the impact of expert selection and elicitation design on elicitation results, a meta-analysis of three recent nuclear expert elicitations was conducted (Anadon et al., 2013) using data from (Abdulla, Azevedo, & Morgan, 2013; Anadon et al., 2012)

. Meta-analysis is a set of statistical techniques used to reconcile and aggregate the results of multiple studies testing similar hypotheses and to thus enhance the overall reliability of findings (Borenstein, Hedges, Higgins, & Rothstein, 2009; Glass, 1976)

It accounts for differences across studies and provides results that are dependent on a consistent set of conditions across observations. This technique has been used in environmental economics since the 1990s (Matarazzo & Nijkamp, 1997; Nelson & Kennedy, 2009), with several recent applications in energy (Barker & Jenkins, 2007; Rose & Dormady, 2011; Zamparini & Reggiani, 2007).

The meta-analysis of three nuclear energy technology elicitations used Individual Participant Data (IPD) meta-analysis as a tool to combine results from multiple expert elicitations into a single data set, facilitating its use in policy decisions and the design of future elicitations. This data was subsequently analyzed to estimate the returns to RD&D investment after controlling for a wide range of observed characteristics at the geographical level, on experts background and elicitation protocol differences. The use of primary data (IPD) is considered the gold standard for systematic reviews because it avoids many of the shortcomings of aggregate meta-analysis: it enables controlling for confounding factors at the individual level and for treatment differences between studies. Moreover, using IPD the study derived results directly and independent of study reporting. This increased the aggregate power of the study, which allowed to more thoroughly scrutinize modeling assumptions (such as the presence of interactions and the linearity of associations) and explore subgroup effects (Borenstein et al., 2009; Gherzi, Berlin, & Askie, 2013; Reade et al., 2009)

It is also important to point out that expert elicitations are used to estimate the distribution of the underlying beliefs held by experts with the largest information sets over an uncertain quantity. Therefore, an expert elicitation study does not rely on asymptotic convergence of sample estimates through the collection of a large number of individual observations, but rather develops

the highest quality representation of the underlying distribution among the most informed experts. In this sense the use of IPD meta-analysis that treats individual experts as single observations relies on a random sampling assumption that the original data collection did not make.

The first objective of this meta-analysis was to understand how public RD&D investment affects experts' 2030 central estimates (50th percentile) and a chosen metric of the uncertainty in assessment (the difference between the 90th and the 10th percentile of expected costs, normalized by the median, $(p90-p10)/p50$). In order to do this, it was important to control for factors affecting the central and uncertainty estimates other than RD&D investments. Thus, the second objective was to understand how research design and expert characteristics affected estimates.

In addition to the public RD&D investment associated with the different expert estimates, the independent variables used to explain the central and uncertainty dependent variables were expert background (industry, academia, and public institution), expert country (American vs. European), technology type (large-scale Gen. III/III+ designs, large-scale Gen. IV designs, and small modular reactor designs), and elicitation mode (in-person vs. online). The relationship between expected costs and RD&D investment was modeled in two ways, representing two different strands of literature concerning the impact of RD&D on innovation. The literature on learning-by-searching (Gruebler, Nakićenović, & Victor, 1999; Junginger, Faaif, & Turkenburg, 2005) often uses a log-log specification to extract a linear relationship between the log of technology costs and the log of RD&D investments, consistent with the observed non-linear scaling in several technology contexts. The literature on diminishing returns to RD&D (Evenson & Kislev, 1976; Hall, Mairesse, & Mohnen, 2009; Popp, 2002) suggests that there are diminishing returns to RD&D investments in some technology areas in which the largest opportunities for breakthrough innovations are exhausted first. Econometrically, this has been represented by introducing a quadratic term to a linear regression, and found to be consistent with the observed relationships in some technology contexts. Estimated negative quadratic coefficients are consistent with diminished returns to RD&D (and are not necessarily inconsistent with learning-by-searching).

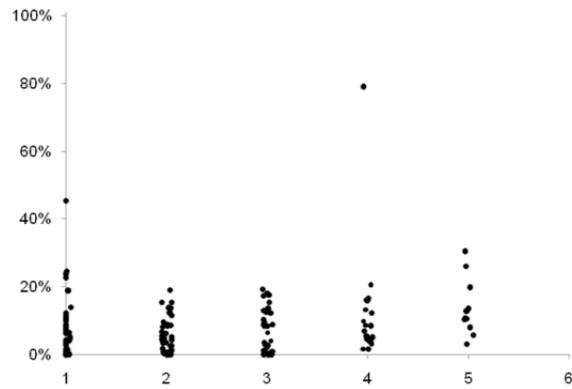
4. Results

In this section we describe the key findings from the elicitation design, modeling, and meta-analysis exercises that provide some insights regarding the use of elicitations on the future of technologies in energy and the role of public RD&D.

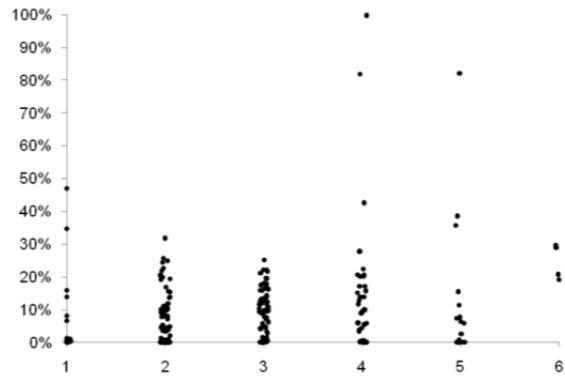
4.1 Including questions about self-rating of expertise

In contrast to widely expressed skepticism about the credibility of technology elicitation in general, US and EU experts did not systematically recommend greater funding levels for the technology areas with which they were most familiar (Figures 1 and 2 show an analysis of U.S. and E.U. experts, respectively). Including a section on self-assessed expertise allowed to assuage some of the concern that experts would be biased towards the sub-technology area that they were most knowledgeable of. Anecdotally, many experts appeared keenly aware of diminishing returns to investment, in part due to near term constraints such as the availability of trained scientists and engineers.

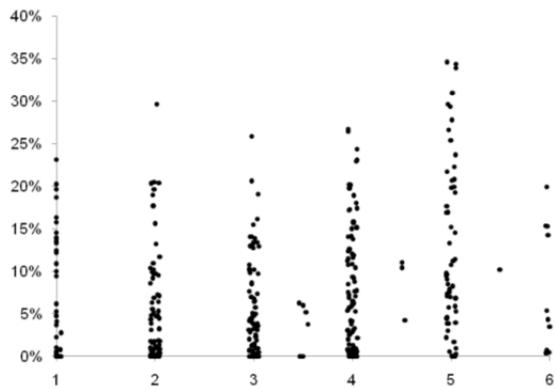
X- axis: Self-rated expertise (1: lowest; 6: highest)
 Y- Axis: Fraction of expert's total investment for a particular technology area



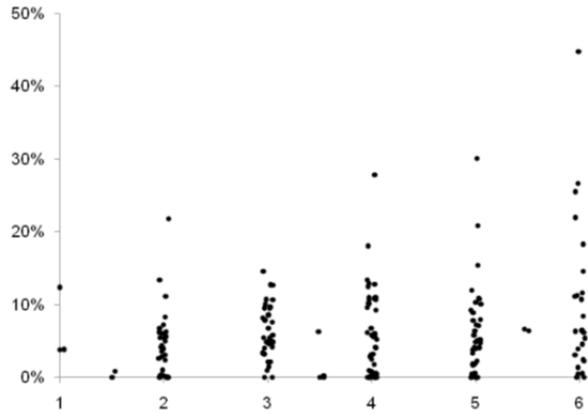
(a) Bioenergy



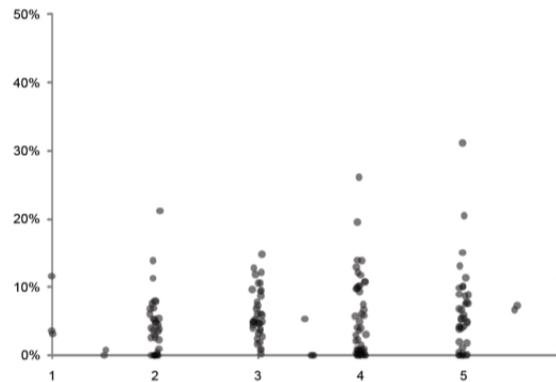
(b) Utility scale energy storage



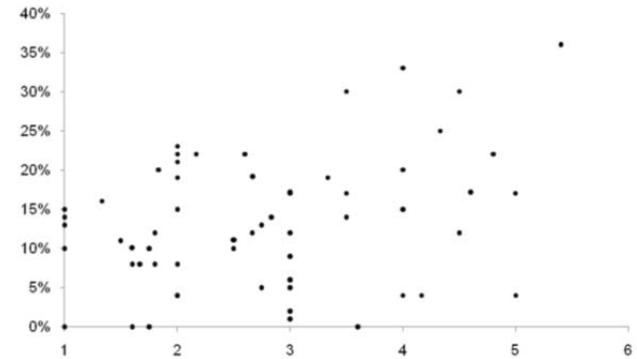
(c) Nuclear energy



(d) Fossil energy and CCS



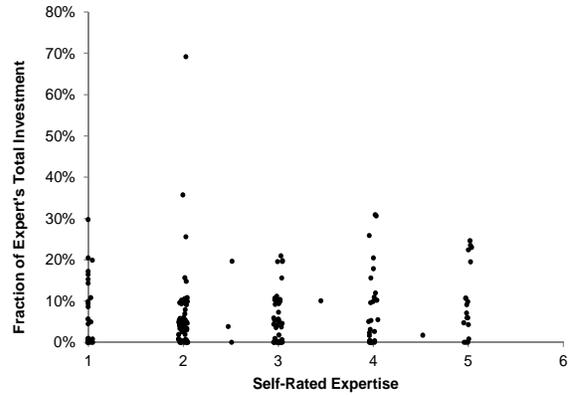
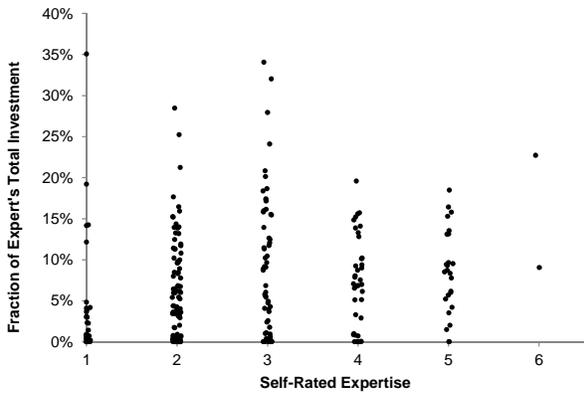
(e) Vehicle technologies



(f) Solar photovoltaics

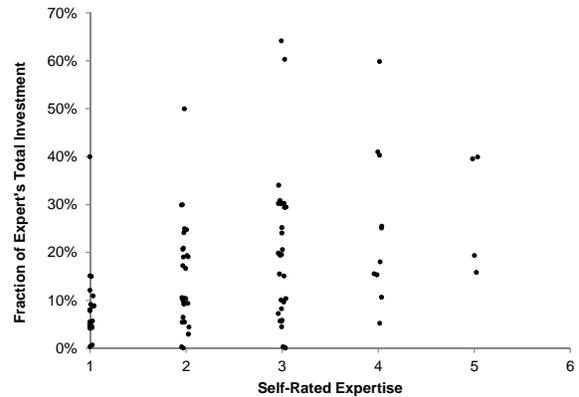
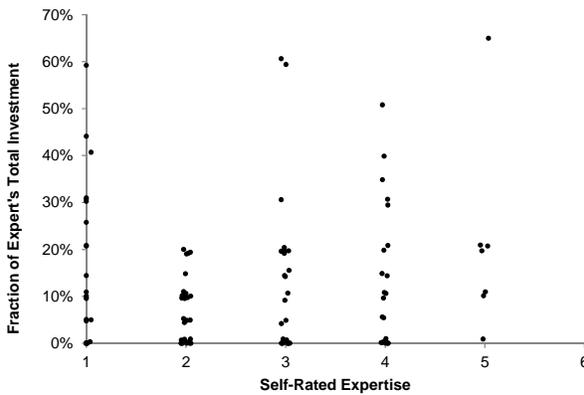
Figure 1: Analysis of expert-recommended budget allocations in areas of self-assessed expertise in Harvard elicitations. The x-axis corresponds to the self-rated expertise (1: I am not familiar with this technology; 6: I am one of the top experts in this technology). The y-axis corresponds to the fraction of the recommended budget that an expert devoted to a particular technology. The graphs represent 6 different elicitations: (a) Bioenergy; (b) Utility scale energy storage; (c) Nuclear energy; (d) Fossil energy and CCS; (e) Vehicle technologies; (f) Solar photovoltaics.

X- axis: Self-rated expertise (1: lowest; 6: highest for nuclear, 1: lowest; 5: highest for all the other technologies)
 Y- Axis: Fraction of expert's total investment for a particular technology area



(a) Nuclear energy

(b) Biofuels



(c) Vehicle technologies

(d) Solar photovoltaics

Figure 2: Analysis of expert-recommended budget allocations in areas of self-assessed expertise in FEEM elicitation. The x-axis corresponds to the self-rated expertise (1: I am not familiar with this technology; 6: I am one of the top experts in this technology). The y-axis corresponds to the fraction of the recommended budget that an expert devoted to a particular technology. The graphs represent 6 different elicitation: (a) Nuclear energy; (b) Biofuels; (c) Vehicle technologies; (c) Solar photovoltaics.

4.2 Conducting elicitation online

Conducting elicitation online and via mail saved time and resources for both researchers and participating experts relative to conducting elicitation in person; conducting elicitation online saved even more time and resources relative to elicitation via mail. A very conservative back of the envelope calculation of the monetary benefits (i.e., excluding benefits in future years,

assuming that researchers travelling to interview experts do not need accommodation, and ignoring the time savings) indicates that online surveys with 11 experts are at least 40% cheaper than in-person elicitation with the same number of experts.¹² However, even though some experts did contact the research team by their own initiative, without any additional information, it is not possible to rule out that the decreased interaction between experts and researchers decreased the value of the information contained in the results, as some experts may have found some of the questions ambiguous (even after extensive pilot testing of the elicitation instruments).

The discussion during the group workshop, which included 18 out of the 60 experts that participated in the FEEM and Harvard nuclear expert elicitation, confirmed that the online tools that provided real-time feedback were useful and that expert interpretation of the questions was consistent with the researchers' intentions. In addition to the qualitative discussion, the robustness of the online elicitation tool was further validated by virtue of very few experts requesting to make changes to their original answers by the end of the workshop.¹³ As explained by the experts, the changes that were made responded to new information made available between the time of the survey and the workshop (the former took place between June 2010 and January 2011, and the latter in April 2012, after the Fukushima nuclear accident in March 11, 2011).¹⁴

¹² The costs of developing the content of in-person and online elicitation instruments are the same. The two main cost differences in using those two types of elicitation instruments stem from the travel cost of flying out researchers to conduct in-person interviews (here we are ignoring the fact that in some cases researchers may have to stay overnight and incur additional accommodation costs), and the cost of hiring a web developer (here we assume, as was the case in the elicitation conducted, that the expertise or time is not available in-house) to build an interactive and easy to use tool. Using the Harvard solar photovoltaics elicitation—which included 11 experts—to conduct a back-of-the-envelope calculation of the monetary savings of conducting the online elicitation we estimated savings of about 50% in the thousands of dollars. Of course, the greater the number of experts participating, the greater the savings, so elicitation contacting more than 11 experts would benefit from even greater savings. This estimate constitutes a low bound given that online tools can be easily used and adapted year after year, which results in additional travel cost savings over time that are not included.

¹³ The workshop was divided into discussion sessions that were design to match the elicitation questions. Each session included a presentation of the results of that part of the elicitation, a moderated group discussion, and a final session in which each expert was provided with a sheet allowing him to privately make changes to his answers to that section (all nuclear experts were men).

¹⁴ Only one U.S. expert revised his Gen. III/III+ BAU cost estimates slightly upward, another U.S. expert revised his Gen. IV cost estimates (also slightly upward), and another U.S. expert revised his BAU SMR cost estimate slightly downward. Two E.U. experts revised their Gen. III/III+ costs slightly (one up and one down), two E.U. experts revised their Gen. IV costs slightly (one up and one down), and one E.U. expert revised his SMR cost estimate downward (Anadon et al., 2012).

We also ran regressions across the Harvard and FEEM elicitations to investigate whether there was any evidence of possible systematic differences in the normalized uncertainty range ((90th-10th)/50th percentile cost estimates of the experts) between online elicitations and in-person or mail elicitations, and across the different RD&D scenarios and time periods (2010 and 2030). Table 1 shows the results of the analysis of the normalized uncertainty range provided by the experts in the Harvard elicitations using dummy variables for online surveys and for different RD&D and technology scenarios. Table 2 shows the regression results for the four in-person FEEM elicitations, which cannot be compared to the 6 Harvard elicitations and the 2 online FEEM elicitations.

Model 1 in Table 1 shows that the normalized uncertainty range in the Harvard data is consistent with experts providing greater normalized uncertainty ranges when conducting the elicitations electronically and online rather than on paper sent by mail. However, we must note that the technology areas are perfectly collinear with the online dummy, which means that further work is needed to disentangle the effect of conducting elicitations online from the differences in normalized uncertainty across technology areas. Model 3 and Model 4 in Table 1 shows respectively that: (a) controlling for unobserved expert-level heterogeneity with expert fixed effects, RD&D scenarios with greater investment than the business-as-usual RD&D scenario had significantly lower normalized uncertainty ranges than the BAU RD&D scenario; and (b) the bioenergy, storage, solar, and nuclear surveys were associated with significantly greater normalized uncertainty ranges than the fossil survey; however, this difference was the smallest for the nuclear; there was not a significant difference in the uncertainty metric between the fossil and vehicles survey.

Turning to Table 2, we find that the analysis of the 4 in-person FEEM elicitations is consistent with the greater RD&D scenarios being associated with increased normalized uncertainty ranges. This result is inconsistent with the Harvard results in Table 1. There are several possible explanations for this difference and unfortunately we cannot conclude what the real cause or causes are. One hypothesis is that fundamentally, U.S. experts believe that more RD&D reduces uncertainty while E.U. experts believe that it decreases it. Another hypothesis is that questions were significantly different and therefore they are not directly comparable. For example, the Harvard survey asked experts to recommend the total amount and specific allocation of RD&D

investments, while the FEEM survey asked experts about increases from the BAU scenario without asking them to design their ideal RD&D program. It is possible that when experts think about their ideal RD&D program they have less uncertainty about the results of their recommendations. It is also possible that when experts think about the impact of RD&D on the aggregated cost of the technologies (as is the case in most of FEEM elicitations) they think about uncertainty differently when compared to components of technology cost (as is the case in most of Harvard elicitations). The difference between the impact of RD&D on the estimates by US and EU experts highlights the complex set of factors involved when making these estimates and are important when considering using the results from different elicitations on similar topics.

Table 1: Analysis of factors associated with differences in normalized uncertainty ranges in the 6 Harvard expert elicitations. The 2030 BAU RD&D scenario and the fossil technology category serve as reference points. $Y = \ln(\text{uncertainty})$.

Variable	Model 1	Model 2	Model 3	Model 4
Online	0.1430** (0.0622)			
2010 BAU		-0.0431 (0.0882)	-0.1140** (0.0504)	-0.0657 (0.0841)
2030 recommended budget		-0.0945 (0.0924)	-0.1055** (0.0479)	-0.0856 (0.0853)
2030 10X recommended budget		-0.0259 (0.0863)	-0.0948** (0.0473)	-0.0168 (0.0800)
Vehicles				-0.1065 (0.1120)
Bioenergy				0.6310*** (0.0964)
Storage				0.7006*** (0.1265)
Nuclear				0.2574*** (0.0669)
Solar PV				0.6894*** (0.0734)
Expert fixed effects	NO	NO	YES	NO
Constant	0.6010***	0.4987***	-1.4974***	0.8255***
R-squared	0.0077	0.0019	0.7419	0.1465
Observations	635	635	635	635

Robust p-values in brackets

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

Notes: The nuclear and solar PV elicitations were conducted online, and the others via mail.

Table 2: Analysis of factors associated with differences in normalized uncertainty ranges in the 4 FEEM in person expert elicitations. The 2030 BAU RD&D scenario and the biofuels technology category serve as reference points. $Y = \ln(\text{uncertainty})$.

	Model a1	Model a2	Model b1	Model b2	Model c1	Model c2
+50% RD&D	0.167* (0.0847)	0.186*** (8.50e-06)	0.165* (0.0868)	0.186*** (8.50e-06)	0.165* (0.0750)	0.186*** (4.86e-06)
+100% RD&D	0.305*** (0.00231)	0.327*** (3.00e-09)	0.298*** (0.00304)	0.327*** (3.00e-09)	0.298*** (0.00244)	0.327*** (2.16e-09)
Solar			-0.110 (0.338)	-0.0267 (0.862)		
Vehicle			-0.151 (0.152)	0.380** (0.0490)		
Cost_CSP					-0.476*** (0.000400)	-0.0267 (0.863)
Cost_EV					-0.104 (0.376)	0.427** (0.0313)
Cost_PHEV					-0.197 (0.105)	0.333* (0.0919)
Cost_PV					0.0429 (0.725)	0.362** (0.0421)
Constant	-0.623*** (0)	-0.439*** (0.00412)	-0.515*** (1.91e-06)	-0.819*** (4.16e-10)	-0.515*** (1.75e-06)	-0.819*** (5.66e-10)
Observations	161	161	161	161	161	161
R-squared	0.058	0.857	0.071	0.857	0.142	0.867
Expert FE	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

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

Note: the “a” and “b” versions of the regression models represent different levels of aggregation in the solar and vehicle technologies.

Overall, the launch, data acquisition, and data processing for the online surveys was faster than the paper surveys. Both groups also learned valuable lessons from the development of their first elicitations (bioenergy energy for Harvard and solar survey for FEEM) that made the development of the remaining elicitations faster. In addition, the results from Table 2 are consistent with online elicitations resulting in greater uncertainty ranges, although further research is needed to demonstrate it. As it will be discussed in section 4.4, the meta-analysis comparing the FEEM and Harvard online nuclear surveys and an in-person nuclear conducted by researchers at Carnegie Mellon University tentatively concluded that conducting the nuclear

elicitation online did not result in statistically different inputs (most notably, that they do not result in greater overconfidence, denoted by smaller uncertainty estimates).

Some evidence of the prospects of using online tools comes from a recent effort unrelated to the researchers in this paper, trying to streamline the use of online elicitations (see NearZero.org for more information) for policy applications.

4.3 Combining elicitations with a group workshop

The Harvard and FEEM groups carried out the same nuclear elicitation in the US and in the EU. The elicitation designed consisted of a two-step procedure of first soliciting experts individually and then following up in a group discussion, which was carried out jointly for the US and EU experts (see Figure 3 for a schematic of the process). From this experience, we identified issues that could arise when each of the two steps is followed as a stand-alone procedure.

As discussed in section 4.2, while cost and performance estimates did not change substantially during the workshop from the individual expert elicitations, the workshop did enrich the information obtained from the elicitations on other topics. We found that the workshop had some impact on the stated RD&D policy objectives that recommended investments were meant to address. Experts who participated in the workshop made some changes (mainly in the form of additions), suggesting that the workshop discussion was helpful in building consensus in this area. RD&D policy objectives that gained priority after the workshop were development of SMRs, risk and safety, and proliferation resistance. EU experts also increased recommended funding for sodium-cooled fast reactors and fuels and materials.

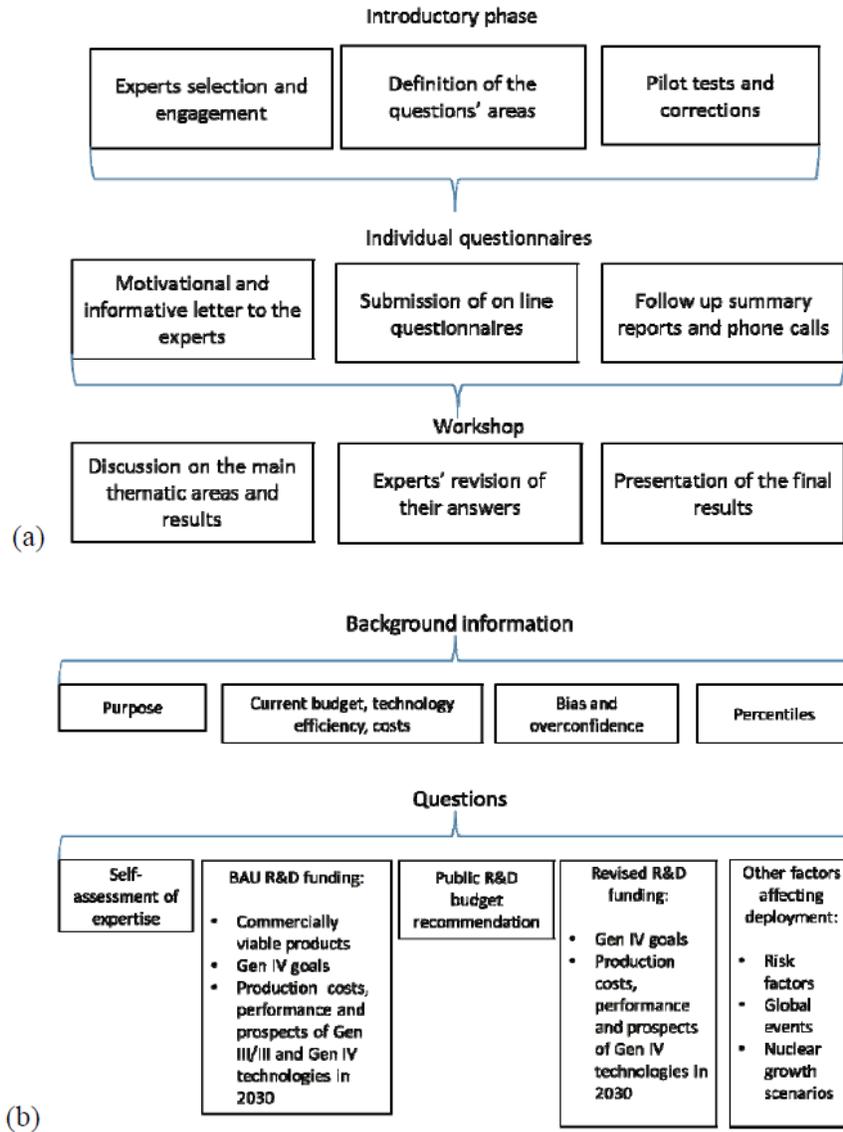


Figure 3: (a) Structure protocol employed in the design of the online elicitation and group discussion; (b) structure of the individual online elicitation instrument (Anadon et al. 2012).

The workshop also resulted in an improved understanding by researchers of how some experts perceived definitional and framing issues that were originally taken for granted. For example, while experts displayed a clear understanding of the questions asked about cost and performance, during the workshop it became clear that different experts were using a different definition of “major radioactivity releases caused by an accident or sabotage.” While some thought that the Fukushima accident would fall under their personal definition of “major radioactivity release,”

others felt that such a description would only apply to a larger accident with more direct casualties. The discussion also revealed that while some experts thought of climate change mitigation as the main goal when making RD&D recommendations, others had multiple goals in mind. This variation in the experts' reasoning would not have been revealed had we pursued only individual elicitation.

The workshop also helped clarify the reasons why U.S. experts placed more emphasis on RD&D to understand fuel cycle economics and reduce fuel cycle costs than E.U. experts and why EU experts thought that it was unlikely that there would be a market for small modular reactors (SMRs) in the future. During the workshop experts explained that fuel cycle costs are a more important issue in the economics of nuclear power in the United States because of its greater focus on private sector involvement, while in Europe such activities might be undertaken by state-owned firms or with other state support. Similarly, experts expressed the belief that financing issues in the United States made SMRs more beneficial in the United States, since they are expected to require smaller lump sum investments.¹⁵

Overall, the combination of the individual online elicitation and expert workshops served to validate the online tool and build consensus in small parts of the survey, while allowing the research team to better understand some of the reasons behind expert answers. The combination of online tools and other tools to increase expert interaction without incurring additional costs is an area of growing interest (Siddharth, Khodyakov, Srinivasan, Straus, & Adams, 2011)

The meta-analysis results also point to new areas to pursue in workshop settings, such as reasons for varying levels of confidence, as well as perceptions about bias due to geography or affiliation type.

4.4 Designing expert elicitation to use as modeling inputs

Even though the Harvard expert elicitation was explicitly designed to provide insights about the optimal allocation and total level of RD&D investments for the 6 technology areas (and 25

¹⁵ Due to the (obviously unplanned) timing of the workshop after the Fukushima disaster, we were also able to determine that the Fukushima disaster did not alter the expert's answers regarding the future of nuclear deployment in the United States and the European Union.

underlying technologies) under consideration, there were some design needs that we were not able to foresee. We now present some of the additional considerations that we identified to improve the elicitation to better match analysis needs.¹⁶ First, obtaining experts estimates of future technology cost over a very large range of RD&D investments, including feasible RD&D ranges well-beyond current levels, can yield additional insights. Experts in the Harvard study were asked to provide estimates of 2030 technology cost and performance under a BAU RD&D funding scenario, their recommended RD&D funding level, and 10 times their recommended funding level. When designing the survey, the researchers believed that this was the maximum feasible range that experts would be able to consider, given that 10-times the recommended levels amounted to 10 to 80 times current funding levels, making assessment of RD&D levels beyond this range suspect to extrapolation bias. The funding levels selected in the Harvard work were sufficient to determine that the current RD&D investment level is too low and that, if properly allocated, \$15 billion in aggregate US RD&D funding could be justified on the basis of aggregate economic benefits. However, because the calculated benefits of RD&D were so large, this range proved too small to estimate the optimal level of RD&D investment. Even though the Harvard study could calculate the rate of decreasing marginal benefits, benefits (in terms of aggregate economic surplus) were still increasing 10% faster than costs at the maximum aggregate range considered, \$15 billion per year (see Figure 4). Other than aggregate economic surplus, there are many other metrics of benefits that one could use (for example, one could use avoided CO₂ emissions for benefits and incorporate opportunity costs for the RD&D costs).

¹⁶ The work by Chan & Anadon (2013) on estimating and optimizing the benefits of energy RD&D portfolios presented here relates to three other pieces of work. Although Blanford (2009) and Davis and Owens (2003) present two frameworks to support investment decisions, they do not justify their assumptions regarding the impact of RD&D on future technology cost and performance, and they do not provide computational flexibility to allow the estimation of optimal RD&D investment levels in a range of technologies at a sufficiently small level of granularity (in the range of millions of dollars) and with the ability to optimize for different goals and risk considerations. Baker & Solak (Baker & Solak, 2011) use elicitation data for three technologies not targeted to inform government investments at the program level and, unlike this work, the R&D investment optimization relies on assumptions about climate damages.

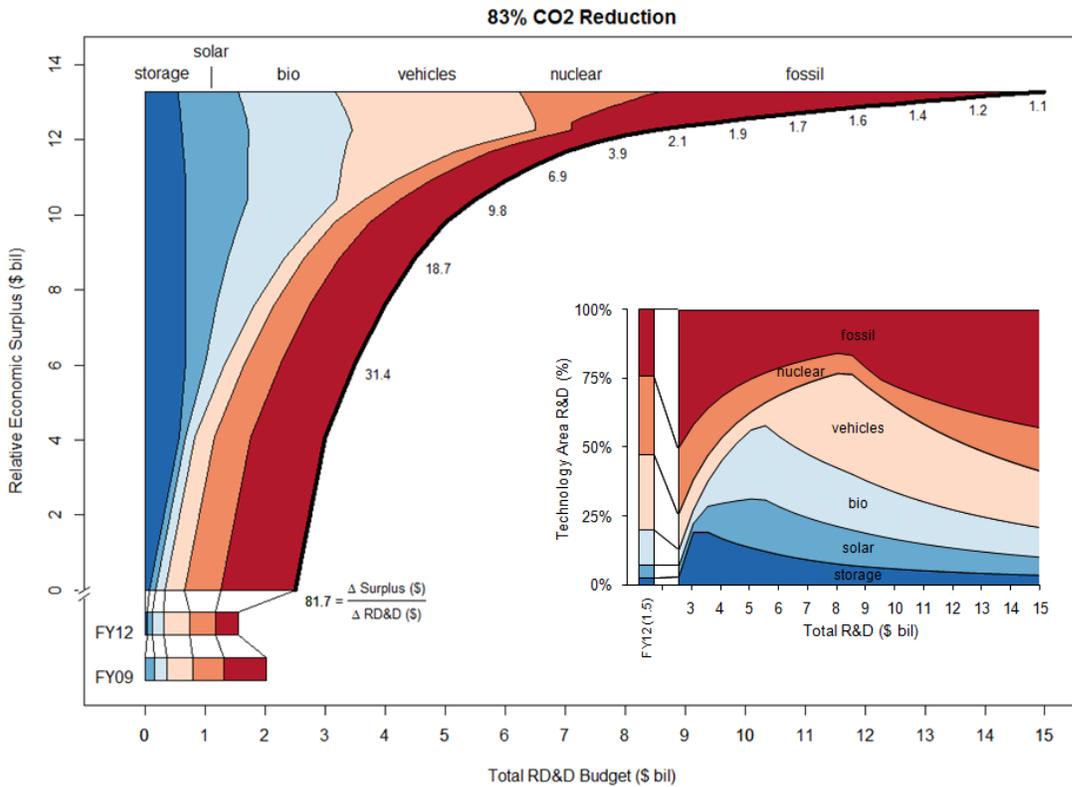


Figure 4: Optimal R&D portfolios under an 83% CO₂ reduction policy. The figure shows the allocation of RD&D funding at different RD&D budget constraints between \$2.5 billion - \$15 billion per year, relative to the Fiscal Year 2009 and 2012 allocations. The dark blue line in the main plots is the maximum expected increase in economic surplus (above the an arbitrary reference point, the expected surplus in the optimal \$2.5bil budget) that can be attained under a given RD&D budget constraint. The small numbers along the black line are estimated marginal returns on investment, calculated by linear approximations to the derivative of the optimal expected surplus at different budgets (Chan & Anadon, 2013)

Second, future elicitation in this area should incorporate questions about the extent to which advances in a particular technology can be assumed to be independent from advances in other related technologies. The Harvard researchers felt that it was reasonable to assume that future advances in some technologies would be uncorrelated with advanced in other technologies (e.g. solar photovoltaics and nuclear technologies). However, due to knowledge spillovers between technology areas, it seemed unreasonable to make this assumption for all technologies (Nemet, 2012). For example, the Harvard bioenergy technology group consisted of technology processes for three products: gasoline substitutes, diesel substitutes, and jet fuel substitutes. Because of the similarity in the technology to produce any of the three products, assuming independence across

the impact of RD&D in the future costs of these technologies did not seem reasonable. The Harvard study also had cases in which complete independence did not seem reasonable across more different technology areas, such as utility-scale energy storage and electric or plug-in-hybrid vehicles. These technologies share similar component technologies (i.e. batteries), but have other very different components. Thus, we developed a table of correlations using our group's expertise in various technology areas (see Table S2 in the SI). To inform future elicitations, we also tested an approach to estimate cross-technology correlations in the vehicles survey. This approach consisted of asking experts to revise their 90th, 10th, and 50th percentile estimates given a particular realization of 2030 costs in a related technology. While we found that most experts were willing and able to think through and answer these questions thoughtfully, including these questions also lengthened an already long elicitation.

Third, we found that asking qualitative questions that asked experts to justify their recommended level of investments and allocation increased our own confidence in the results and their external credibility. This is something that we did not include in the first elicitation (the bioenergy elicitation), but that we included in the subsequent five elicitations. For more information on what some of these qualitative questions focused on, the reader can access the links to the nuclear survey in the SI of (Anadon et al., 2012).

Fourth, the large number of experts included in the Harvard elicitation (more than 100), required substantial preprocessing before summary results could be presented. We developed an importance sampling technique to reduce the computational requirements of assessing the RD&D allocations and forecasts of many different experts. However, for the parsimony of presenting results, we eventually had to either select or aggregate experts. Anadon et al. (2011) relied on three "expert scenarios", labeled, "optimistic," "middle," and "pessimistic." Each expert scenario grouped the answers of the 6 most optimistic, central, and pessimistic experts. As shown in Figure 4, even increasing RD&D investments from a BAU budget of \$2 billion to \$82 billion/year, and utilizing assumptions from the most optimistic experts, CO₂ emissions are not expected to decrease substantially from current levels. Thus, creating "expert scenarios" allowed researchers to calculate high and low bounds of benefit metrics that did not depend on the choice of expert.

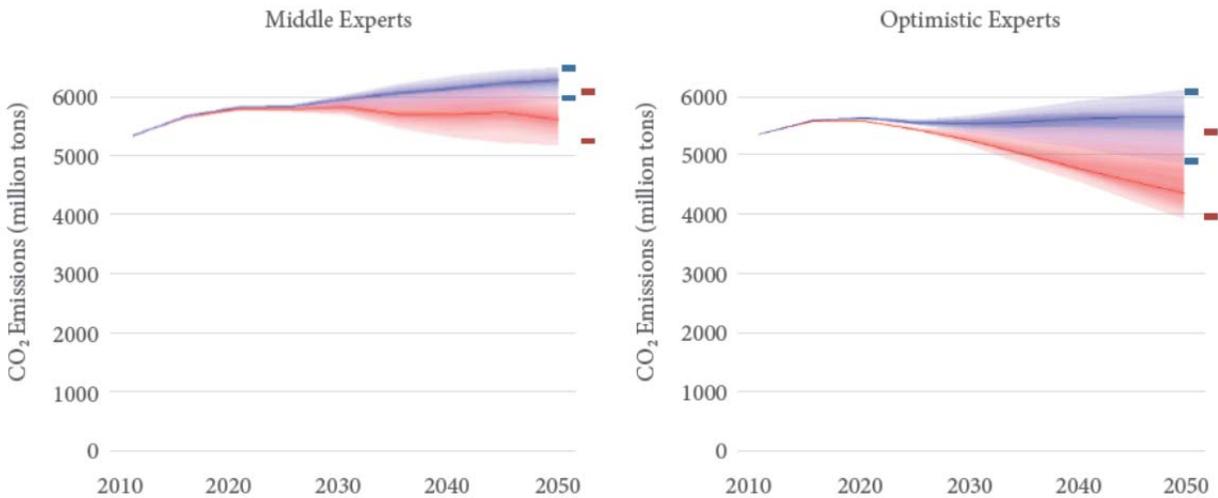


Figure 4: U.S. energy-related CO₂ emissions under (a) business-as-usual federal energy RD&D investment and no additional demand-side policies (blue) and (b) ten times the experts’ average recommended federal energy RD&D investments (somewhere between \$49 and \$82 billion/year) (red), with no additional demand-side policies, using “middle of the road” and “optimistic” experts’ technology cost projections. Note that optimistic experts were optimistic about technological progress in general, and not necessarily optimistic about the effects of RD&D (Anadon et al., 2011b).

4.5 Using meta-analysis to improve elicitation usability and design

The meta-analysis of the nuclear elicitations evaluated the impact of expert selection (background and country) and elicitation design (technology granularity and online vs. in person mode). The goal of this exercise was to inform future elicitations and to better capture the experts’ thinking on the impact of public nuclear RD&D on future technology costs for modeling and policy analysis.

The results of the two non-linear models we specified, log-log and linear-quadratic, were consistent in terms of the statistical significance and sign of the estimated effects. Here, we discuss the key insights from the log-log model of the experts’ central estimate of nuclear power overnight capital cost in 2030 (which are inspired by the relationship between cost and RD&D investment put forward in the learning-by-searching framework, as described in section 3.2).

(Anadon et al., 2013) shows the quantitatively large influence of expert composition on the range of expert inputs available for policy analysis. Controlling for expert affiliation, expert country of origin, and technology type, the coefficient of the RD&D variable increases by 25% relative to

the estimated coefficient in the reduced form model (namely, RD&D on costs). On average, we find that a doubling of the yearly public nuclear RD&D budget in the U.S. and the E.U. is associated with an 8% decrease in nuclear costs in 2030, *ceteribus paribus*. We also found that experts from public institutions have estimates of overnight capital costs that are about 14% higher on average than those of academics and that estimates from industry experts are even higher, on average around 31% higher than academics. Expected overnight capital costs are approximately 22% lower for experts in the USA when compared to experts in the European Union. Technology type also is a statistically significant determinant of 2030 costs: overnight capital costs are expected to be higher for both Gen. IV and SMR technologies with respect to Gen. III/III+ technologies by roughly 23% and 24%, respectively.¹⁷ With regards to the impact of RD&D and other explanatory variables on our uncertainty metric of choice (the 90th percentile estimate less the 10th percentile estimate, normed by the 50th percentile estimate), we find that higher or lower levels of investment are not systematically associated with narrower or wider uncertainty ranges. However, U.S. experts have around 16% wider uncertainty ranges compared to EU experts. The uncertainty range for SMRs is about 14% smaller than that for large scale Gen. III/III+, suggesting that experts are relatively more confident about their cost estimates for these systems. This was a somewhat surprising finding considering that SMRs are expected to be delivered to the site fully constructed from the manufacturing facilities, yet current experience is limited and no operating licenses have been issued in the United States or the EU.

In ongoing work, Anadon, Nemet, and Verdolini are conducting a larger meta-analysis that includes expert elicitations in other energy technology areas—solar PV, coal with carbon capture and storage, and biofuels—to determine the extent to which the impact of different variables is consistent across different technology areas. The increased variation among these studies, as well as the increase observations, will enable more precise estimation of both expert and elicitation design effects and will allow to gauge differences in experts' assumptions about the returns of RD&D in different technological areas.

¹⁷ The Anadon, Nemet & Verdolini (2013) study found that the in-person variable (accounting for the observations obtained through an in-person interview as opposed to through an online tool) becomes negative and significant when expert fixed effects are included, although it is difficult to draw conclusions about this effect since it requires inclusion of unobserved expert characteristics for it to become significant. This tentative result is consistent with results in Table 1 in this paper, but the tentative nature of this analysis requires that the in-person effects be a focus of future work assembling additional elicitation data ensuring that more than the 3% of observations are in-person.

5. Conclusions and future work

The findings presented in this paper provide lessons for the future design and use of expert elicitations to inform policy decisions on public RD&D investments. The lessons from this work are applicable not only to energy, but also to other technology areas that receive substantial government RD&D support, such as health, [defense?] and agriculture. The findings presented in this paper stem from several pieces of work: (1) 10 expert elicitation exercises encompassing 6 energy technology areas each and conducted between 2009 and 2011 by Harvard researchers and FEEM researchers, respectively (Anadon et al., 2011a; Anadon et al., 2012; Bosetti et al., 2012; Catenacci M. et al., 2013; Chan et al., 2011; Chan et al., 2011; Fiorese et al., 2013); (2) a paper that combined online elicitations on nuclear energy from a research group at Harvard and at FEEM (Anadon et al., 2012); (3) a paper that relied on the Harvard elicitations to estimate the optimal RD&D allocation across six technology areas using the MARKAL energy-economic model combined with an importance sampling and optimization technique (Chan & Anadon, 2013); and (4) a meta-analysis of nuclear expert elicitations from FEEM, Harvard, and Carnegie Mellon University conducted by researchers at FEEM, Harvard, and the University of Madison-Wisconsin to obtain insights about the impact of expert selection and survey design on the impact of RD&D on technological change (Anadon et al., 2013). Below we summarize five key findings outlined in section 4.

First, we find that mail and online expert elicitation tools can be used to obtain expert elicitation estimates more cost-effectively than in-person interviews without compromising the quality of expert answers. This finding relies on insights from the expert workshop that followed FEEM and Harvard nuclear elicitations. This finding is also conditional on appropriate preparatory work by the eliciting research team that included extensive background research on the topic, pilot testing the elicitation instrument, including background material that discussed biases and confidence, and the utilization of numerous interactive visual aids. Conducting elicitations online in particular can contribute to an easier institutionalization of the process.

Second, we find that asking experts to self-assess their level of expertise in specific technologies and processes, to justify their RD&D priorities, and to identify non-RD&D-related factors that would affect the future of specific technologies, increases both the researchers' confidence in the

level of intellectual engagement of the experts and in the external credibility of the results. For example, experts were not systematically recommending larger amounts of funding to their areas of expertise, providing some evidence that experts were at least not solely motivated by self-interest to receive more funding. Other studies have used similar assessments of experts' expertise to weight responses, but in this work we did not use them to maintain the full dispersion in outcomes and because we wanted to remove issues related to the reliability and comparability of self-assessments of expertise in the modeling effort.

Third, to support decisions about RD&D investments in different technology programs, it can be useful to push experts to consider a wide range of scenarios, including scenarios at the boundary of their private information set, to explore potentially-desirable scenarios far from current activities without undue extrapolation bias. In addition, elicitations should include questions to allow the deduction of correlations across technology improvements. Alternatively, researchers (or analysts) can create a separate elicitation targeting correlations.

Fourth, some important policy insights can be derived by creating scenarios without aggregating experts. Insights regarding the need to put in place additional policies beyond RD&D investments to meet CO₂ emissions reductions goals, and the decreasing marginal returns to RD&D investments, were independent of whether or not modeling included experts that were optimistic, central, or pessimistic regarding forecasted 2030 technology costs.

And fifth, expert selection has a large and very significant impact on elicitation results, indicating that experts from the private sector, academia, and public institutions, as well as experts from different countries, have different private information sets and beliefs. An elicitation exercise that sought to include all perspectives would need to include experts from all of these backgrounds. Further, the meta-analysis exercise allowed researchers to better understand estimates of the impact of RD&D on technology costs.

The insights regarding expert elicitation design and utilization to support energy RD&D investment decisions are not only applicable to decisions on public RD&D investments in the energy area. Public RD&D investments in other sectors also face questions regarding the extent to which they should be guided purely by scientific merit or by mission. For example, there have been calls to increase the extent to which funding in the R&D budget of the National Institutes of

Health (NIH) should consider disease burdens (see review by (Sampat, 2012)). This approach would require not only that greater fractions of the NIH budget be allocated to specific diseases, but also some consideration of the extent to which additional research could result in improvements. Industrial research institutions could also implement some of the insights and methods discussed in this paper, as they also deal with investing in projects with uncertain returns that will only impact their bottom line if they are diffused in the market.

Although the combination of insights from this body of work improves our confidence in the use of expert elicitations to inform RD&D decisions in the energy sector and (we would argue) beyond, there are several avenues for future research that could further improve our understanding. As it has already been mentioned in this paper, some of this work is ongoing. Future research could randomize experts into three different groups to complete the same elicitation in-person, online, or via mail, respectively to conduct a more systematic evaluation of whether there are any systematic differences in the results. Additional meta-analysis work including elicitations for energy technologies beyond nuclear energy would establish the extent to which expert background and country variables change across technologies and the extent to which returns to RD&D vary across different technologies—which can be a powerful modeling tool going forward. Ongoing work involving three major teams involved with energy economic models (GCAM at the Pacific Northwest National Laboratory, WITCH at FEEM, and MARKAL at Brookhaven National Laboratory) is using aggregates of elicitation results from different studies using equal weights among the three major studies from the University of Massachusetts Amherst, FEEM, and Harvard University. Finally, the question of whether or not to aggregate expert answers to model future technical change and the uncertainty around it was not a focus of this work (the focus was on insights robust to different “expert scenarios”). Identifying the benefits of aggregating expert assessments may ultimately require ex-post analysis of previous elicitations against the realized technical change.

Acknowledgments

Laura Diaz Anadon and Gabriel Chan acknowledge funding from the Climate Change Initiative of the Doris Duke Charitable Foundation, and a grant from BP-International Limited on Energy, Climate & Security Policy. Valentina Bosetti and Elena Verdolini received funding from the

European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 240895 – project ICARUS “Innovation for Climate Change Mitigation: a Study of energy R&D, its Uncertain Effectiveness and Spillovers.” Gregory Nemet acknowledges support from the Wisconsin Alumni Research Foundation.

References

- Abdulla, A., Azevedo, I. L., & Morgan, M. G. (2013). Expert elicitation of the cost of small modular reactors. *Proceedings of the National Academy of Sciences*, *110*(24), 9686–9691.
- American Energy Innovation Council. (2010). *A business plan for america's energy future. full report*. (No. June). Washington D.C., United States: American Energy Innovation Council.
- Anadon, L. D., Bunn, M., Chan, G., Chan, M., Jones, C., Kempener, R., Narayanamurti, V. (2011a). *Transforming U.S. energy innovation*. (). Cambridge, MA, United States: Energy Technology Innovation Policy Group, Belfer Center for Science and International Affairs, John F. Kennedy School of Government, Harvard University.
- Anadon, L. D., Bunn, M., Chan, G., Chan, M., Jones, C., Kempener, R., . . . Narayanamurti, V. (2011b). *Transforming U.S. energy innovation*. (No. November). Belfer Center for Science and International Affairs, Harvard Kennedy School, Cambridge, MA, United States: Energy Technology Innovation Policy research group. (To be published by Cambridge University Press in 2013).
- Anadon, L. D., Bosetti, V., Bunn, M., Catenacci, M., & Lee, A. (2012). Expert judgments about RD&D and the future of nuclear energy. *Environmental Science & Technology*, *46*, 11497-11504.

- Anadon, L. D., Nemet, G. F., & Verdolini, E. (2013). The future costs of nuclear power using multiple expert elicitations: Effects of RD&D and elicitation design. *Environmental Research Letters*, 8, 034020.
- Anadon, L. D. (2012). Missions-oriented RD&D institutions in energy between 2000 and 2010: A comparative analysis of China, the United Kingdom, and the United States. *Research Policy*, 41(10), 1742–1756.
- Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson (Ed.), *The rate and direction of inventive activity* (pp. 609-626). Princeton, NJ, USA: Princeton University Press.
- Baker, E., Chon, H., & Keisler, J. (2009a). Advanced solar R&D: Combining economic analysis with expert elicitations to inform climate policy. *Energy Economics*, 31, 537-549.
- Baker, E., & Solak, S. (2011). Climate change and optimal energy technology R&D. *European Journal of Operations Research*, 213, 442-454.
- Baker, E., Chon, H., & Keisler, J., M. (2008). *Advanced nuclear power: Combining economic analysis with expert elicitations to inform climate policy*. ().Social Science Research Network.
- Baker, E., Chon, H., & Keisler, J. (2009b). Advanced solar R&D: Combining economic analysis with expert elicitations to inform climate policy. *Energy Economics*, 31, S37-S49; S37-S49.

- Barker, T., & Jenkins, K. (2007). *The costs of avoiding dangerous climate change: Estimates derived from a meta-analysis of the literature.* (). New York, NY, United States: United Nations Human Development Report.
- Blanford, G. J. (2009). R&D investment strategy for climate change. *Energy Economics*, *31*(1), S27-S36.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, U.K.: John Wiley & Sons, Ltd.
- Bosetti, V., Catenacci, M., Fiorese, G., & Verdolini, E. (2012). The future prospect of PV and CSP solar technologies: An expert elicitation survey. *Energy Policy*, *49*, 308-317.
- Catenacci M., Verdolini E., Bosetti B., & Fiorese G. (2013). Going electric: Expert survey on the future of battery technologies for electric vehicles. *Energy Policy*, *In Press*, <http://www.sciencedirect.com/science/article/pii/S0301421513005855>.
- Chan, G., Anadon, L. D., Chan, M., & Lee, A. (2011). Expert elicitation of cost, performance, and RD&D budgets for coal power with CCS. *Energy Procedia*, *4*, 2685-2692.
- Chan, G., & Anadon, L. D. (2013). Utilizing expert assessments to inform allocating government energy RD&D investment portfolios. *Submitted*,
- Cooke, R. L. (1991). *Experts in uncertainty: Opinion and subjective probability in science.* New York City, NY, USA: Oxford University Press.

- Curtright, A. E., Morgan, M. G., & Keith, D. W. (2008). Expert assessments of future photovoltaic technologies. *Environmental Science & Technology*, 42(24), 9031-9038.
- Davis, G. A., & Owens, B. (2003). Optimizing the level of renewable electric R&D using real options analysis. *Energy Policy*, 31, 1589-1608.
- EERA. (2010). *8th Framework Programme (FP8) position paper of the European Energy Research Alliance (EERA)*. October. Brussels, Belgium: European Commission.
- EIA. (2011). *Direct federal financial interventions and subsidies in energy in fiscal year 2010*. (). Washington D.C.: U.S. Energy Information Administration, U.S. Department of Energy.
- European Commission. (2007). *Communication from the commission to the council, the European Parliament, the European Economic and Social Committee, and the Committee of the Regions "limiting global climate change to 2 degrees Celsius the way ahead for 2020 and beyond"* .. Brussels, Belgium: European Commission, COM(2007) 2 final.
- Evenson, R. E., & Kislev, Y. (1976). A stochastic model of applied research. *The Journal of Political Economy*, 84(2), 265-281.
- Fiorese, G., Catenacci, M., Verdolini, E., & Bosetti, V. (2013). Advanced biofuels: Future perspectives from an expert elicitation survey. *Energy Policy*, 56, 293–311.
- Fishbone, L. G., & Abilock, H. (1981). MARKAL, a linear-programming model for energy systems analysis: Technical description of the BNL version. *International Journal of Energy Research*, 5(4), 353-375.

- Gallagher, K. S., Anadon, L. D., Kempener, R., & Wilson, C. (2011). Trends in investments in global energy research, development, and demonstration. *Wiley Interdisciplinary Reviews: Climate Change*, 2(3), 373-396.
- Ghersi, D., Berlin, J. A., & Askie, L. (2013). *Cochrane Collaboration—Prospective Meta-Analysis Methods Group*. <http://pma.cochrane.org>: The Cochrane Collaboration.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Gruebler, A., Nakićenović, N., & Victor, D. G. (1999). Dynamics of energy technologies and global change. *Energy Policy*, 27, 247-280.
- Hall, B. H., Mairesse, J., & Mohnen, P. (2009). Measuring the returns to R&D. *National Bureau of Economic Research Working Paper Series, No. 15622*
- Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision*. Chichester, West Sussex, United Kingdom; New York City, NY, USA: Wiley.
- IEA. (2013). *International Energy Agency Energy R&D Statistics*. (No. Statistics. Available at: <http://www.iea.org/stats/rd.asp>). Paris, France: International Energy Agency.
- IEA, OPEC, OECD, & World Bank. (2010). *Analysis of the scope of energy subsidies and SUGgestions for th eG-20 initiative. joint report*. (). Toronto, Canada: International Energy Agency, Organization fo the Petroleum Exporting Countries, Organization for Economic Cooperation and Development, and The World Bank.

- Jenni, K. E., Baker, E. D., & Nemet, G. F. (2013). Expert elicitations of energy penalties for carbon capture technologies. *International Journal of Greenhouse Gas Control*, 12, 136-145.
- Junginger, M., Faaif, A., & Turkenburg, W. C. (2005). Global experience curves for wind farms. *Energy Policy*, 33, 133-150.
- Keeney, R. L., & Winterfeldt, D. (1991). Eliciting probabilities from experts in complex technical problems. *Transactions on Engineering Management*, 38, 191-201.
- Matarazzo, B., & Nijkamp, P. (1997). Meta-analysis for comparative environmental case studies: Methodological issues. *International Journal of Social Economics*, 24(7/8/9), 799-811.
- McNerney, J., Farmer, J. D., & Trancik, J. E. (2011). Historical costs of coal-fired electricity and implications for the future. *Energy Policy*, 39, 3042-3054.
- Meyer, M. A., & Booker, J. M. (1991). *Eliciting and analysing expert judgment: A practical guide*. London, U.K.: Academic Press Ltd.
- Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.
- Narayanamurti, V., Diaz Anadon, L., Breetz, H., Bunn, M., Lee, H., & Mielke, E. (2011). *Transforming the energy economy: Options for accelerating the commercialization of advanced energy technologies*. Cambridge, MA, United States: Energy Technology Innovation Policy group. Belfer Center for Science and International Affairs.

- NCEP. (2004). *Ending the Energy stalemate. A bipartisan strategy to meet America's Energy Challenges.*. Washington D.C., United States: National Commission on Energy Policy.
- NCEP. (2007). *Energy policy recommendations to the President and the 110th Congress.*.The National Commission on Energy Policy.
- Nelson, J., & Kennedy, P. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental & Resource Economics*, 42(3), 345-377.
- Nemet, G. F. (2012). Inter-technology knowledge spillovers for energy technologies. *Energy Economics*, 34(5), 1259-1270.
- Nemet, G. F. (2013). Technological change and climate-change policy. In J. Shogren (Ed.), *Encyclopedia of energy, natural resource and environmental economics* (Amsterdam, The Netherlands ed., pp. 107-116) Elsevier.
- Nemet, G. F., & Kammen, D. M. (2007). U.S. energy research and development: Declining investment, increasing need, and the feasibility of expansion. *Energy Policy*, 35(1), 746-755.
- NRC. (2007). *Prospective evaluation of applied energy research and development at DOE (phase two)*. National Academies Press, Washington D.C., United States: National Research Council.

- PCAST. (1997). *Federal energy research and development for the challenges of the twenty-first century*. (No. Chapter 4). Washington D.C., United States: President's Council of Advisors on Science and Technology, Executive Office of the President.
- PCAST. (2010). *Report to the president on accelerating the pace of change in energy technologies through an integrated Federal energy policy*. ().President's Council of Advisors on Science and Technology. Executive Office of the President.
- Popp, D. (2002). Induced innovation and energy prices. *American Economic Review*, 92, 160-180.
- Pugh, G., Clarke, L., Marlay, R., Kyle, P., Wise, M., McJeon, H., & Chan, G. (2011). Energy R&D portfolio analysis based on climate mitigation. *Energy Economics*, 33, 634-643.
- Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Oxford, UK: Addison-Wesley.
- Reade, M. C., Delaney, A., Bailey, M. J., Harrison, D. A., Yealy, D. M., Jones, P. J., . . . Angus, D. C. (2009). Prospective meta-analysis using individual patient data in intensive care medicine. *Intensive Care Medicine*, 36(1), 1-21.
- Rose, A., & Dormady, N. (2011). A meta-analysis of the economic impacts of climate change policy in the united states. *The Energy Journal*, 32(2), 143-166.
- Sampat, B. N. (2012). Mission-oriented biomedical research at the NIH. *Research Policy*, 41, 1729-1741.

Schock, R. N., Fulkerson, W., Brown, M. L., San Martin, D., Greene, R. L., & Edmonds, J.

(1999). How much is energy research & development worth as insurance? *Annual Review of Energy and Environment*, 24, 487-512.

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

Zamparini, L., & Reggiani, A. (2007). Meta-analysis and the value of travel time savings: A transatlantic perspective in passenger transport. *Networks & Spatial Economics*, 7(4), 377-396.

Supplementary Information

Table S1: Descriptive statistics of Harvard expert elicitation data for analysis on the impact of uncertainty (636 observations, of which 283 were online).

Variable	Description	Mean	Standard Deviation	Minimum	Maximum
Uncertainty	(P90-P10)/P50 estimates by expert	0.7972	0.8271	0.2	9.7
2010 BAU	RD&D scenario	0.25	0.4332	0	1
2030 BAU	RD&D scenario	0.25	0.4332	0	1
2030 recommended budget	RD&D scenario	0.25	0.4332	0	1
2030 10X recommended budget	RD&D scenario	0.25	0.4332	0	1
Fossil	Coal, coal with CCS, gas, gas with CCS (4 types)	0.1905	0.3929	0	1
Vehicles	Advanced, conventional, hybrid, plug in hybrid, electric, and hydrogen (6 types)	0.1978	0.3985	0	1
Bioenergy	Gasoline, diesel, and jet fuel substitutes (3 types)	0.0879	0.2833	0	1
Storage	Utility scale energy storage in the hour range	0.1282	0.3345	0	1
Nuclear	Large-scale Gen. III/III+, Large-scale Gen IV, SMR (3 types)	0.2747	0.4466	0	1
Solar PV	Residential, commercial, or utility (3 types)	0.1209	0.3261	0	1
Online	Whether or not survey was administered online	0.3956	0.4892	0	1

Table S2: Correlation matrix

advanced internal combustion engine vehicles; HYB: light-duty hybrid vehicles; PEV: light-duty plug-in hybrid vehicles; FCV: light-duty fuel cell vehicles; THR: large-scale Gen. III/III+ nuclear power; FOR: large-scale Gen. IV nuclear power; MOD: small and medium factory-built nuclear power; PVR: residential photovoltaic solar power; PVC: commercial photovoltaic power; PVU: utility photovoltaic power.