Expert views - and disagreements - about the potential of energy technology R&D

Laura Diaz Anadon[†] · **Erin Baker**[†] ·
Valentina Bosetti[†] · **Lara Aleluia Reis**

Abstract Mitigating climate change will require innovation in energy technologies. Policy makers are faced with the question of how to promote this innovation, and whether to focus on a few technologies or to spread their bets. We present results on the extent to which public R&D might shape the future cost of energy technologies by 2030. We bring together three major expert elicitation efforts carried out by researchers at UMass Amherst, Harvard, and FEEM, covering nuclear, solar, Carbon Capture and Storage (CCS), bioelectricity, and biofuels. The results show experts believe that there will be cost reductions resulting from R&D and report median cost reductions around 20% for most of the technologies at the R&D budgets considered. Although the improvements associated to solar and CCS R&D show some promise, the lack of consensus across studies, and the larger magnitude of the R&D investment involved in these technologies, calls for caution when defining what technologies would benefit the most from additional public R&D. In order to make R&D funding decisions to meet particular goals, such as mitigating climate

[†] All authors contributed equally.

Erin Baker
Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Amherst, MA, USA
Tel.: 413-545-0670
Fax: 413-545-9265
E-mail: edbaker@ecs.umass.edu

Laura Diaz Anadon
Harvard Kennedy School, Harvard University, Cambridge, MA, USA

Valentina Bosetti
Department of Economics, Bocconi University, Milano, Italy
Fondazione Eni Enrico Mattei, Milano, Italy
Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy

Lara Aleluia Reis
Fondazione Eni Enrico Mattei, Milano, Italy
Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy

change or improving energy security, or to estimate the social returns to R&D, policy makers need to combine the information provided in this study on cost reduction potentials with an analysis of the macroeconomic implications of these technological changes. We conclude with recommendations for future directions on energy expert elicitations.

Keywords R&D investments · energy technology · expert elicitation · risk and reward · low-carbon innovation

1 Introduction

Mitigating climate change will require innovation in energy technologies. There is agreement that both government policies to limit carbon dioxide emissions and R&D investments will be necessary [27,21] to spur the innovation needed to mitigate climate change at reasonable cost [24]. Policy makers are faced with the question of how to promote this innovation, and whether to focus on a few technologies or to spread their bets. A key difficulty in making decisions about public R&D investments is the uncertainty about the impacts of those investments. In addition, budgetary constraints in governments create real tensions between focusing on making significant progress in a small number of technologies as opposed to experimenting with a wider set of options. Moreover, questions on how best to support R&D investment decisions when multiple areas compete for funds are common to fields other than energy, including health, agriculture, and defense.

In the field of energy, the U.S. National Research Council [33] has explicitly recommended that the U.S. Department of Energy uses probabilistic assessment based on expert elicitations of R&D programs in making funding decisions. A key benefit of probabilistic expert elicitations over other approaches, such as those used in roadmaps (e.g. IRENA 2015 [26]), is that projections in roadmaps and other modelling exercises are usually deterministic and not transparent (the source of the estimate is unclear). Conversely, elicitations allow for a more transparent methodology that crucially can account for the deep uncertainty, in our example related to technical change. A number of studies of expert judgments on single technologies have been published in many places in the literature by a range of researchers [36,17,13,44,11,1,29,40,43,4]. However, due to widely differing assumptions and methods it has been very difficult for decision makers to get a sense of what conclusions can be drawn from the sum of these publications and reports [14]. This study is the first attempt to collect, harmonize, and analyze the insights that can be derived from a large number of expert judgment studies performed by different research teams on different energy technologies. We examine to what degree the international community is coming to consensus around the prospects of key energy technologies that have been highlighted as crucial for climate mitigation [25] and the role that public R&D can play. This rigorous and transparent methodology for analyzing the results of multiple elicitations side by side may be used

to inform decisions in other fields where scattered knowledge on the impact of R&D investments exists.

We bring together three major expert elicitation efforts carried out by researchers at UMass Amherst, Harvard, and FEEM, [8,7,5,6,3,2,12,18,19, 11], covering nuclear, solar, Carbon Capture and Storage (CCS), bioelectricity, and biofuels, and aimed at evaluating the impact that public R&D investments in the United States and the European Union could have on the cost of these technologies. All the above mentioned studies were carried out by means of expert elicitations, a structured process for eliciting subjective probabilities from experts [23]. Following the literature, each study provided significant materials to minimize biases.

It should be noted that the three studies were not designed to be combined or compared. This is both the motivation for, and a challenge to, the present analysis. The UMass Amherst, FEEM, and Harvard studies were funded, respectively, by the Department of Energy of the US Government, the EU 7th Framework Program, and Doris Duke Charitable Foundation, respectively. Each study could be taken alone and used to inform R&D decision making. However, using a single study would not represent the best information available, given that we do not have an understanding of whether any of the studies or data points within the studies provides more predictive information or a better representation of uncertainty. Thus, we develop and implement a methodology aimed at supporting decision makers that represents the extent of this uncertainty. We note that this study is not intended to be a meta-analysis evaluating the impact that the expert selection process or other factors in the study design might have had on the median estimates or the uncertainty range (see [4] and [41] for this approach), but rather a presentation of the current state of information in as clear, transparent, and comparable way as possible, given the challenges.

The differences between the surveys allow us to consider a variety of assumptions and detect whether there are robust insights or common patterns on the relative effectiveness and riskiness of R&D programs in these different technologies to be drawn by these exercises taken together. We set out to answer the following question: is there consensus regarding the technologies that would be most affected by R&D investment; or, at least in terms of potential cost reductions, is the best strategy to keep options open by investing widely and gathering more information on the potential evolution of costs?

Each expert elicitation survey covers one technology area and provides information on multiple experts' views from multiple sectors (academia, industry, and other public institutes) on the prospects for future costs conditional on specific R&D funding levels. Funding levels are grouped into three broad categories, Low (which is consistent with a business-as-usual (BAU) scenario for FEEM, an increase of 50% to 200% over BAU for Harvard, small investments, independent from the BAU, into specific technologies for UMass), Medium (ranging between an additional 50% to a 16-fold increase over low) and High (ranging between an additional 30% to a 10-fold increase over medium). While both Harvard and FEEM included demonstration expenditures, UMass asked

questions about smaller R&D scenarios that did not include demonstration expenditures. Given these different approaches, the absolute values of the funding levels vary across studies. The intention of this work is to summarize the full spectrum of information available in these three groups of studies, since it is not possible to determine which of the approaches used by the various studies is more accurate. On the one hand, the differences across elicitations make comparisons precarious. On the other hand, the diversity allows us to get a range of results that is wider than what could be obtained with one single study. Aggregating multiple studies allow us to include more expert views, in this case spanning the United States and EU; while having different question formats, metrics, and styles, covers elicitation format uncertainty. While the use of different R&D levels makes it difficult to fully interpret the differences between studies, it can also help alleviate possible errors arising from anchoring on a particular R&D level. From our results is appears that R&D levels may bias results, for example leading to over- or under- confidence for low or high levels. Thus, pulling together elicitations spanning various parts of the R&D range may provide a more complete picture of the uncertainty involved. Important future work would include a controlled study investigating how to best calibrate experts to different R&D levels. Most importantly, policymakers interested in the results of expert elicitations need a way to compare results across studies, even if the studies are not controlled and fully comparable [14]. Elicitations are expensive and time consuming, so policy makers (understandably) would like to use all available information. One approach is to simply tell policy makers that the comparison cannot be made, and allow them to make their own guesses as to how to bring together different pieces of information. This may lead to picking of one single study, effectively narrowing the uncertainty that entering the decision making process. Instead we have performed the comparison, using transparent methods and making caveats as clear as possible on the data challenges.

Our analysis is based on a total of 114 experts' views (see appendix for a complete list) collected through expert elicitation (in person, via mail, or online) in a time span of three years (2009—2011). The appendix reports common and heterogeneous assumptions of different surveys on R&D scenarios, funding amounts (see Table 8 of Section 7 of the appendix), elicited metrics and protocols, and how the surveys were harmonized to produce common metrics at common target dates. Experts took between 2—5 hours to complete the elicitations. These were difficult exercises that required them to hold policies and materials prices constant, while conditioning on R&D levels. While this is challenging, it is not unprecedented: the approach is similar to that in an elicitation of global mean temperature in which experts were asked about the role of additional research in reducing uncertainty [42]. The particular design of the elicitations we compare stemmed from the need by policy makers to understand and isolate the role of R&D. While such challenging exercises tend to lead to large uncertainty ranges (as it did in this set of elicitations), we argue that these ranges in fact represent the state of knowledge on the impact of R&D, and therefore are useful.

2 Materials and Methods

The number of experts included in each study varied from 3, for the UMass solar study, to 31, for the FEEM nuclear study. The average number of experts per study was just under 13, slightly larger than a typical Expert Elicitation study [39]. All 14 studies had experts from academia; 11 studies had experts from government or NGOs; and 10 studies had experts from industry. Studies varied, with UMass having over 60% from academia and only 10% from industry; FEEM had over 50% from government and NGOs, the remainder divided evenly between academia and industry; Harvard had 33% from government, and with 35%, had the largest percentage from industry. See Table 4 in the SI for a summary by technology.

Both Harvard and FEEM surveys collected information on self-reported expertise. However, we note that there is no evidence that weighting by self-reported expertise significantly changes or improves reporting [10]. Moreover, while some meta-analyses have found differences in assessments based on background [4,41], these differences are not consistent across studies of different technologies (e.g., in the two citations provided, private sector experts were consistently more optimistic than academics in the future of solar technologies but more pessimistic than academics for nuclear), nor is there any evidence on which category of expert is more likely to be more accurate. Thus, this variation in background is another strong reason for presenting all the data available.

Experts had given their consent, either in written [3,2,12,18,19,11] or verbal [8,7,5,6] form, and have agreed on having their names published. However, the data is reported anonymously thus it is never possible to link data with individual experts. A total of 165 surveys/interviews with experts were completed by the three teams. The details on the methods used by the three teams are presented in the separate papers. Some experts participated in multiple surveys, and some survey results were not used in this study due to missing data. The elicitation protocols contained background information and training about biases. Experts were asked about the future cost of technology under various R&D scenarios. The elicitations varied in their approaches, some asking for percentiles, some for probabilities, and some for both. An example of what the elicitations look like is available in the links provided in the Supporting Information of Anadon et al (2012) [2].[1] As discussed in the appendix, a harmonization and aggregation protocol was developed to allow the expert estimates from different elicitations to be compared alongside each other. This protocol involved converting data into comparable currencies, as well as converting data to have consistent metrics and years. After the three studies were harmonized, we aggregated them using a linear opinion poll with equal weights. Thus, the "combined" results represent a pooling of the three studies

---

[1] Link to U.S. nuclear elicitation survey: `https://erd3.cloudapp.net/nuclear_energy` Username: erd3 Password: laura Link to E.U. nuclear elicitation survey: `https://erd3.cloudapp.net/icarus` Username: icarus Password: laura .

which we always present in parallel to the individual study data for the sake of transparency.

3 Results

We first look at cost reduction potentials associated to R&D, with a formulation that accounts for possible negative improvements. We define the improvement ratio, $R_{i,j}$, as the improvement in technology metric i when R&D funding increases from $j-1$ to j

$$R_{i,j} = \frac{M_{i,j-1} - M_{i,j}}{M_{i,j-1}} \tag{1}$$

j represents the funding level of Low, Medium, or High, and $j-1$ is the next lower funding level; $M_{i,j}$ is the value of the metric for technology i, for funding level j. For example, $M_{Solar,medium}$ would give the value of a specific measurement of the Levelized Cost Of Electricity (LCOE) for solar in 2030, say the median, at *Medium funding*. Note that our definition of the improvement ratio does not capture opportunity costs from investing the funds in energy technologies instead of elsewhere [34]. The improvement ratio must be put in the context of the cost of R&D and the social benefit from R&D for a full analysis of the value of different R&D portfolios.

In order to calculate the improvement ratios from increasing R&D funding levels, shown in Figure 1, it is necessary to make an assumption about the correlation of outcomes resulting from different funding levels. We focus on the results under one extreme assumption: the outcomes from different funding levels are perfectly correlated. This assumption would hold if (1) higher funding levels always fund the same projects as lower funding levels, plus additional projects; and (2) the additional projects are perfectly correlated with the initial projects. The first argument explicitly holds for the UMass studies. It does not explicitly hold for FEEM and Harvard, where the experts may have been thinking about different projects at different levels of funding. The second condition seems unlikely to hold, but may be an approximation if we believe that there are underlying factors that impact success in an entire technology category. In the appendix we consider the implications of a second, extreme assumption: that the outcomes from different funding levels are independent. We find that the central insights regarding the median expected improvement ratio and decreasing impacts as R&D increases remain the same. The main differences in these results are the greater uncertainty that one expects when one assumes complete independence, and the fact that there are many negative values. The negative tails are much larger than the positive tails since improvements (cost decreases) are bounded by 100%, whereas cost increases are unbounded. It seems unlikely that increasing R&D will in fact lead to significantly worse outcomes, thus we argue that the assumption of perfect correlation is closer to reality than independence.
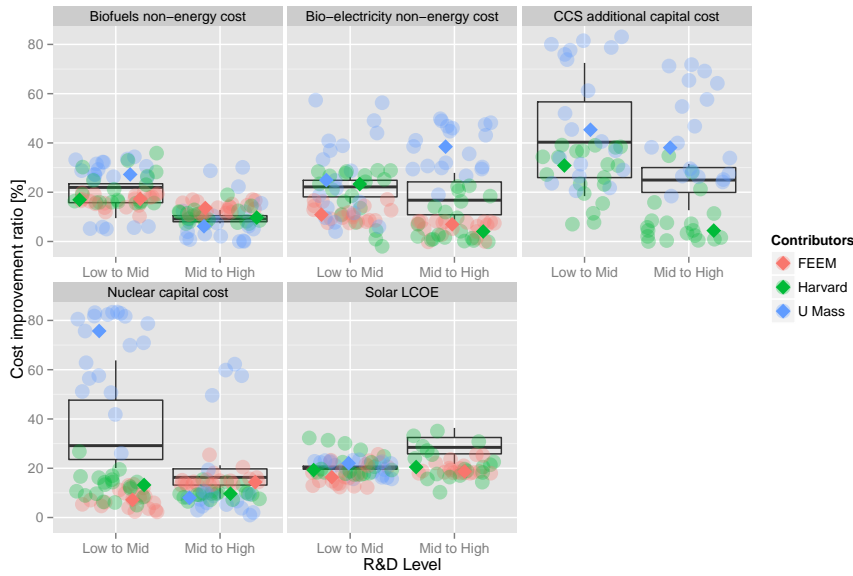
Fig. 1 Cost Reductions from increasing public R&D funding from low-to-mid R&D levels and from mid-to-high R&D levels, assuming improvements from increments of R&D are perfectly correlated. The box plot represents the 5th, 25th, 50th, 75th and 95th percentiles of the distribution aggregated across teams (the combined analysis) the dots represent samples (in increments of 5 percentage points from the 5th to the 95th percentile) from the FEEM (red), Harvard (green), and UMass (blue) studies. The solid diamond-shaped points refer to the median improvement ratios.

Figure 1 shows the distribution of the relative improvement ratios to R&D, $R_{i,j}$, moving from the low-to-mid R&D (left) and from the mid-to-high R&D (right) for cost metrics for five technologies (LCOE for solar ($/kWh); non-energy cost for bio-electricity ($/kWh) and biofuels ($/gge), which considers both thermochemical and biochemical conversion pathways; additional capital cost for CCS in coal power plants ($/kW); and overnight capital cost for large-scale nuclear power ($/kW), which includes both Gen. III/III+ and Gen. IV systems). For each of the 5 metrics presented, we plot the results of sampling from each of the aggregate distributions derived using data from the FEEM, Harvard and UMass surveys. The solid diamonds depict the 50th percentile for each of the three studies. We also include the box plot of the combined distribution, which weights equally the results of the three studies. Each individual point represents an equally likely point from a particular study. For example, consider the results clustered around the "low-to-mid" vertical bar in the upper left chart. Each pale blue circle represents an equally likely improvement ratio for biofuels non-energy cost from the UMass data. Specifically, each circle is calculated using the ith percentile costs under the mid and low scenarios, where i ranges by 5's from 5 to 95. The results are presented as a "cloud" – the dispersion around the x-axis is to allow visualization of all

the results; it does not have any meaning. The box plots are calculated from the ith percentiles of the aggregated distribution across the three elicitation teams; they do not represent the distribution of the individual importance ratios. Therefore, in cases such as nuclear for which the distribution is bimodal (with UMass results being more optimistic), the aggregate distribution falls in the middle. This result emphasizes the need to depict the individual expert, the various study aggregates, and the overall aggregate results to increase the transparency in the study and to avoid giving the false impression of consensus. Here we summarize the findings.

Prospects for technologies in the combined analysis: The median improvement ratios to R&D tend to be somewhat similar across the technologies, with all technologies except CCS having median improvement ratio around 20% for the low-to-mid R&D scenario (low-to-mid). CCS has a median improvement ratio of 40% for low-to-mid, and has the highest improvement ratio for that level of funding in both studies that considered it. In the low-to-mid category, Nuclear follows CCS as the highest improvement ratio, while in the mid-to-high category, solar and CCS have the highest improvement ratio. In order to put the numbers shown in Figure 1 in perspective with other related work, we want to note that there are some similarities between the concept of improvement ratio from our work and the empirical literature estimating historical "Learning by Searching" (LBS) rates for various technologies. The LBS literature, which reports rates of cost reductions per doubling of R&D investments, are of similar magnitude to those found in this forward looking paper—e.g., 12.6% in wind [31], 5.3% for thermal solar power, 8.9% for combined heat and power, 20.6% for small hydropower, and 23.8% for nuclear power (light water reactors) [28]. It should be emphasized, however, that a direct comparison between the average 20% improvement ratio that emerges from Figure 1 and the LBS estimates above is not fully appropriate, for two key reasons. First, the coefficients in the LBS literature refer to a doubling of R&D levels, while in this paper the improvement ratios in Figure 1 show changes from one fixed R&D level to another. Thus, these scenarios do not map exactly to doublings. Second, the LBS literature estimates the LBS rate using historical data and usually relies on a two factor model, which includes LBS and learning by doing (LBD); thus they do not explicitly separate out other factors, such as material prices or economies of scale [35]. The forward looking elicitation studies we use in this paper tried to specifically isolate the role of public R&D from all these other factors by asking experts to condition on public R&D levels alone. Similarly, it is important to note that the values shown in Figure 1 cannot be compared to the literature estimating historical returns to R&D [22], as these papers measure the returns to society rather than improvements in terms of cost reduction.

While we cannot validate the accuracy of predicted outcomes in 2030, it is interesting to compare these values, where possible, to observed prices in 2014. We focus on the changes seen in the Harvard study between 2010 and the low R&D scenario in 2030, as Harvard was the only study to include a 2010 baseline

(with the exception of the FEEM nuclear study). We limit our comparison to solar PV and bioelectricity because there have not been enough nuclear plant, large-scale CCS, or large-scale second generation biofuel constructions in the regions of interest to allow meaningful comparisons for those three technologies. We use LCOE data from the International Renewable Energy Agency (IRENA, 2015) showing the cost evolution for those two technologies between 2010 and 2014.

The LCOE of solar PV decreased by 37% between 2010 and 2014, compared to the Harvard average prediction of a 65% reduction between 2010 and 2030 under a BAU R&D scenario. These two values are relatively consistent in terms of the trend implied over the past 4 years; thus the elicitations appear to be relatively consistent with observations. Moreover, the observed 37% decrease was due almost entirely to materials prices and economies of scale in manufacturing (with only 3.7% attributed to technology advances). These factors were not included in the elicitations, since R&D decision makers have little influence over them. But this implies that (1) when elicitation results are combined with models, these other factors should be included; and (2) it may be useful to incorporate empirical data such as experience curves into future elicitation studies.

For the case of bioelectricity, the LCOE between 2010 and 2014 has slightly increased, compared to the Harvard average prediction of a 19% decrease between 2010 and 2030. In both solar PV and bioelectricity, when looking at all the studies, a significant increase in R&D spending from "low" to "mid" was expected to yield additional cost reductions of 20% on average.

Agreement among studies: Figure 1 also highlights the differences between the aggregate distributions of the potential improvement ratio among studies. UMass estimates, marked by the blue circles, are generally more optimistic than FEEM and Harvard, with the exception of solar. We also note a greater agreement among the three studies for the improvement ratio on the cost of biofuels and solar PV. This could be due to the fact that there is more recent experience building PV panels and biorefineries than building CCS, nuclear and novel bioelectricity plants.

Focusing on the median improvement ratios, we find that different technologies do better in the surveys of the different teams, with each of the 5 technologies showing up in 1st or 2nd place in one of the teams, and only CCS showing up twice in the top 2, as shown in Table 6 of Appendix. We note that while biofuels is one of the technologies with the largest improvement ratios for FEEM, it is among the lowest improvement ratios for the other two teams. This may be related to the prominent investment in biofuels in the U.S. by the DOE and BP that was already taking place at the time of the elicitations, which did not have an analogue in the EU [3,19]. This may have led U.S. experts to predict a smaller impact for any additional funding. The teams also result in different ranges, with UMass having the largest range among the improvement ratios, and FEEM the lowest.

In general, we stress that the disagreement about the impact of R&D on the future cost of technologies in the different studies may result from different factors. First, the technologies themselves may be at an early stage so that there has not yet been enough time for the creation and circulation of a common knowledge pool. It is possible that different experts have a different experiences with and knowledge of various aspects of the technology, which could result in information being compartmentalized and not widely available. Second, some of the estimates obtained might be more reliable than others, but the science of expert elicitations at this point does not allow us to discriminate among the existing data points. This disagreement can be addressed by more technology R&D and by more research into the study of expert elicitations. For example, the different studies asked questions at different aggregation levels, ranging from specific components (e.g., module cost, inverter, etc.) to highly aggregated metrics, such as LCOE. While LCOE is much simpler to ask and leads to shorter surveys, it requires a great deal of implicit calculations and estimations by the experts. Future work could be aimed systematically determining the impacts of this survey design choice. Certainly this analysis calls for retrospective research evaluating the extent to which different elicitations were more or less predictive.

Agreement among experts and differences between the European and US experts. For the FEEM and Harvard studies we decompose uncertainty into two parts: the contribution to the aggregated variance of individual uncertainty and between-expert uncertainty (see Table 7 section 6 in the Appendix). We find that for most of the studies over half of the uncertainty is due to disagreements between experts. This may indicate that individual experts are over-confident (which has been shown to be important in the literature on biases and decision making under uncertainty [38] and/or that information about the technologies is not well-diffused through the community. Of interest is the fact that European experts in the FEEM solar study are much closer to consensus than the US experts in the Harvard study. This may reflect differences in the respective solar communities, or it may reflect key differences in the two elicitation formats. A validation of the different formats including experts from multiple countries would allow to tackle this important research question.

Decreasing marginal improvement ratios to R&D investment. One of the proposed models in the literature of endogenous growth theory [30] is that of decreasing returns to R&D, following a fishing-out model. By plotting improvement ratios versus R&D efforts we see the emergence of decreasing marginal improvements in almost all of the technologies across the surveys (see Figure 2) along with the wide dispersion of additional R&D investment covered by the three elicitation teams. It is useful to compare the range over which R&D investments exhibit decreasing marginal improvement ratios in Figure 2 to current observed investments. For example, in 2014 U.S. public investments by the Department of Energy were 232 million for bioenergy, 257 million for

solar (which includes other technologies), 493 million for nuclear energy, and 435 million for coal and carbon sequestration [20]. While the three elicitation teams all have different amounts defining "low" R&D investment (see Table 8 in Section 7 in the online appendix), most of the observed US investments fall within the range that are associated with decreasing marginal improvements according to the results from the elicitations. The exceptions are: nuclear in the FEEM and Harvard studies, and CCS in the Harvard studies; the "low" R&D amounts in these studies are higher than the observed, thus we cannot conclude they are in the range of decreasing marginal improvement ratios. Here we note that we are considering "returns" only in terms of a cost reduction in the technology; we are not considering a reduction in the risk of climate damages or an improvement to other societal impacts, such as health damages from air pollution. Additional modelling is needed to consider the impact of R&D on climate or other risks (e.g., [9,3], etc.).
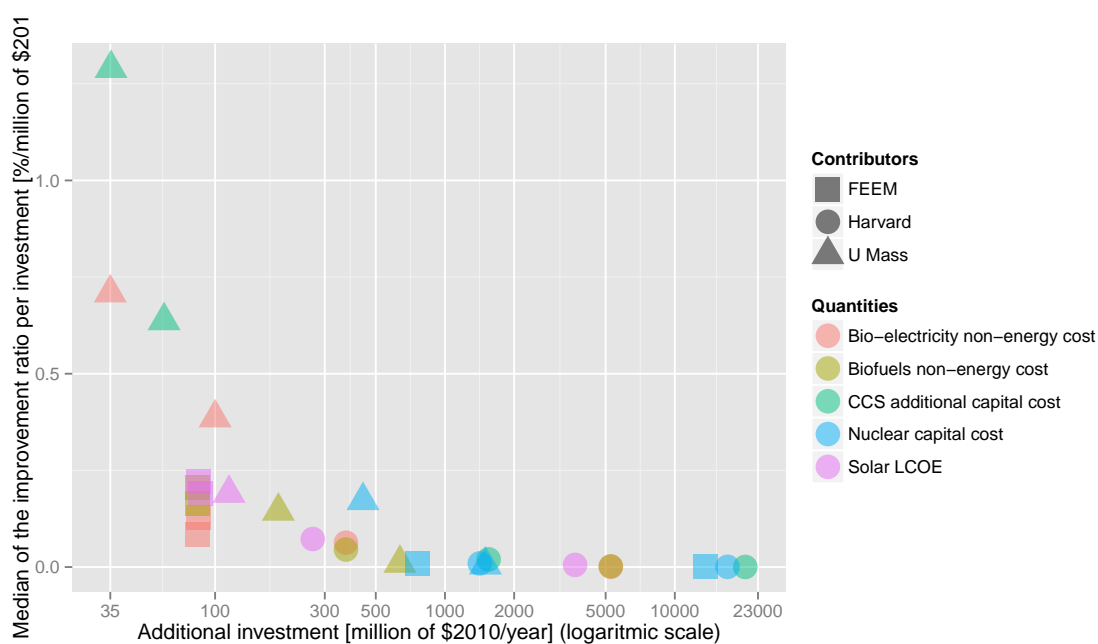
Fig. 2 Median improvement ratio divided by additional R&D investment level for different technologies and teams. The x-axis shows the additional R&D investment in a logarithmic scale [million of $2010/year].

Uncertainty of Improvement Ratios. One way to think about uncertainty over the outcome of R&D programs is to consider the range of possible outcomes, especially in the context of the median outcome. When looked at this way

in Figure 1, solar is the least risky, since there is very little variation in the aggregate improvement ratio within each and between the three elicitation teams. However, another, more rigorous way, to consider uncertainty is through the idea of stochastic dominance.

Figure 3 shows the Cumulative Distribution Functions (CDFs) for the improvement ratios distributions in Figure 1. From these we can see stochastic dominance relations. A distribution First Order Stochastically Dominates (FOSD) another if all decision makers (who prefer more to less) prefer it; visually, a distribution FOSD another if its CDF lies entirely to the right. A distribution Second Order Stochastically (SOSD) dominates another if all risk averters prefer it; visually, a distribution X SOSD Y if the area below the CDF for X measured from left to right, is less than the area below the CDF for Y measured from left to right. For example, the solid green line in the upper left panel of Figure 3 is showing that, in the combined analysis, the improvement ratio from low to mid investment has a 50% chance (from the y axis) of being 40% or less (from the x-axis at bottom of figure); it has a 75% chance of being less than 55%. We can see that the least risky investments are solar mid-to-high (dashed orange line), nuclear low-to-mid (continuous purple line) and CCS low-to-mid (continuous green line), since these lines are the lowest and farthest to the right. These three technologies FOSD all other distributions (excepting a slight crossover for CCS with solar low-to-mid).Note, however, that the sizes of the investments (represented by the thickness of the line) into these three technologies is higher than those into biofuels, bio-electricity, and solar at the low-to-mid level.

Including in the discussion all four panels, the most risky investments are all found in the mid-to-high funding level, including biofuels and nuclear, which are FOSD by all the other distributions, and bio-electricity, which is SOSD by all other distributions. In addition, nuclear, contrarily to the combined case, does not dominate in either the FEEM and Harvard results. Solar and CCS are close to second order stochastically dominating in two of the studies each (FEEM and Harvard for solar and Harvard and UMass for CCS), indicating that the prospects for R&D induced improvements in those two technologies may be more robust than for others. Solar, at both investment levels, has very vertical CDFs across the three studies, meaning it has a tight range with little variance. While the medians for most of the improvement ratios tend to be fairly close - between 16-30% for most technologies - the upside (the high percentiles) varies significantly — between 21-64% for the same set of technologies. In looking at the individual team panels, we see that all the teams follow the same patterns in riskiness, in terms of stochastic dominance, that they did in the medians, except Harvard bio-electricity, which shows no stochastic dominance over any of the technologies, despite having the 2nd highest median.

In order to best support decision-making, we show both the pooled results of the three studies (in the upper left panel); and the results of each individual study. This is useful in many ways. For near-term decision making, the gold standard is to use the best available information; thus the combined data

provides a starting point to support decisions about R&D investments by analyzing the impact of those investments on future costs. However, where there is considerable disagreement or wide uncertainty across the various studies, as in this case, it is important that decision makers focus on near term decisions that maximize option value — that is, near term decisions that increase the flexibility to respond to new information in the future.
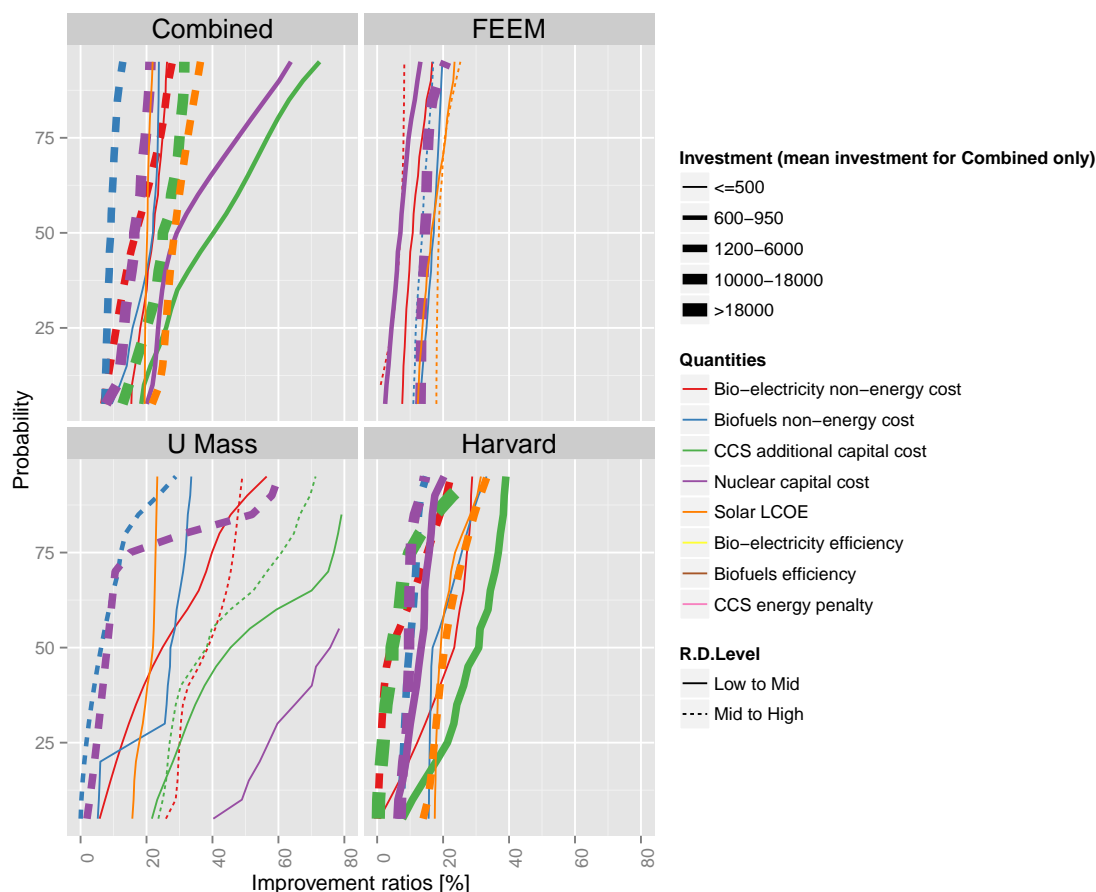
Fig. 3 Cumulative distribution functions for the improvement ratios distributions in Figure 1 by study, technology, and increase in R&D investment. The thickness of the lines denotes the additional R&D investment level associated with that technology investment. For the combined panel R&D investments are mean investments across studies.

4 Conclusions

We have consolidated what the scientific community across both sides of the Atlantic believe will be the impact of public R&D investments on the 2030 costs of five key energy technologies. Going back to our original question, we find that scientific knowledge on the impact of R&D investments on the future of energy technologies does not justify a strategy of focusing on one or two technologies: no single technology stands out across the studies as having consistently higher improvements or being less risky. While this analysis should be complemented by energy economic models to support decision making, the wide range of uncertainty regarding the impact of R&D suggests it is best to focus on the option value of investments, whether into R&D or into expert elicitation, and that it is too early to pick winners. The finding of decreasing marginal improvement ratios to investment in any one category further reinforces this conclusion, as increasing funding significantly for only one technology in the short term is expected to result in relatively small benefits. There is, however, some evidence to suggest that experts generally believe that the greatest risks are in large investments in nuclear, biofuel, and bio-electricity technologies; and that solar has relatively low risk at low investment costs.

This analysis provides a unique overview, including the largest number of top scientific experts to date, regarding the expected impact of public R&D investments on technological change in key energy technologies. The results from the harmonization of the three separate expert elicitation studies are important, because each of the three studies, taken alone, indicates that some technologies are more promising than others. Yet, when taken together, the results are not as clear. The process outlined in this paper allows us to analyze all the evidence available; the results of this process point to the necessity of maximizing flexibility, rather than focusing on a specific technology. Further research — into the technologies themselves and into methods for assessing the future prospects of the technologies, such as expert elicitations — should be pursued before committing too strongly to one path over another. The reason for this is that different expert elicitations point to different technologies having greater potential for cost reduction. Thus, while harmonizing and combining disparate studies is a considerable challenge, it provides a great deal of value over and above the individual elicitation studies, which are themselves very labor and resource intensive.

It is important to note that in order to make R&D funding decisions to meet particular goals, such as mitigating climate change or improving energy security, or to estimate the social returns to R&D, policy makers need to combine the information on cost reduction potentials with an analysis of the macroeconomic implications of these technological changes. Small changes in the cost of key technologies may have a disproportional impact on the cost of mitigating climate change [37,32]; whether from possible threshold effects, post-R&D improvements, or broader social benefits. Similarly, the costs of the R&D programs, including the opportunity costs, must be included in any analysis. Thus, information on the potential for success alone, although a key

element of such analysis, is not enough to fully identify a portfolio of investments to contribute towards a particular societal goal.

Based on our analysis, we suggest directions for improvements in future elicitation studies. First, it would be useful to incorporate experience curves and estimates of external factors into future elicitations.

Second, future work should experimentally test experts' ability to calibrate to different R&D budgets. For example, a study could create groups of experts spanning a range of characteristics and conduct the same elicitation differing only in R&D levels. If the results from the study with the small R&D budgets continue to be systematically more optimistic than those conditional to much larger budgets, this could suggest that when thinking about small projects, experts focus on specific programs they know well and tend to be overconfident, and/or that when thinking in aggregate, experts tend to anchor more to the status quo.

Third, future work may more fully develop methods for discriminating among the reliability of different experts' and studies' estimates. There is a significant body of work showing that the calibration of experts' judgments on test questions that are related to the questions of interest are good predictors of their accuracy on the questions of interest (this is referred to as the classical method, see [15] for a review). The elicitations reviewed here did not ask test questions; thus, we cannot use this method to discriminate among studies or experts. However, the reliability of test questions and their ability to discriminate in favor of those experts who will perform best in real-life elicitations is subject to a lively debate (see [10,15]). The elicitations reviewed here did not ask test questions; thus, we cannot use this method to discriminate among studies or experts. In the present context, the fact that near term test questions would be used to calibrate experts' judgments over events happening in the far future further complicates the construct of a testable hypothesis to settle the issue. However, introducing these test questions in future elicitations will enable retrospective research to assess this open question.

Fourth it is important that studies make expert elicitation results available in both raw (as coming from the elicitations) and standardized forms (such as what was done in the present paper), as well as provide information about experts' calibration, where available. As more information becomes available, this would help to empirically validate subjective probabilities in the energy field, similarly to what has been done in other areas ([16,15]). Crucially, having disparate studies focused on the same topics will help elucidate what approach yields more accurate results. The ideal process would be to ex-ante standardize multiple studies on the same technology, using similar target years and amounts of R&D funding; however, research groups tend to work in a de-centralized manner performing elicitations independently, thus this may be very hard to achieve. A key compromise is that all elicitation studies are fully transparent about all assumptions, including currency years, target years, R&D expenditures, and any other external conditions.

And fifth, future work could be aimed at systematically determining the impacts of survey design choice, specifically related to the level of aggregation.

The different studies asked questions at different aggregation levels, ranging from specific components (e.g., module cost, inverter, etc.) to highly aggregated metrics, such as LCOE. While LCOE is much simpler to ask and leads to shorter surveys, it requires a great deal of implicit calculations and estimations by the experts.

References

1. Abdulla, A., Azevedo, I., Morgan, M.G.: Expert assessments of the cost of light water small modular reactor. PNAS **110.**, 39,686–9691 (2013)
2. Anadon, L.D., Bosetti, V., Bunn, M., Catenacci, M., Lee, A.: Expert judgments about rd&d and the future of nuclear energy. Environ. Sci. Technol. **46**, 11,497–504 (2012)
3. Anadon, L.D., Chan, G., Lee, A.: Transforming U.S. Energy Innovation, chap. Expanding and Better Targeting U.S. Investment in Energy Innovation: An Analytical Approach. Cambridge University Press, Cambridge, UK (2014)
4. Anadon, L.D., Nemet, G., Verdolini, E.: The future costs of nuclear power using multiple expert elicitations: effects of rd&d and elicitation design. Environmental Research Letters **8**, pp. 034,020 (2013)
5. Baker, E., Chon, H., Keisler, J.: Advanced solar r&d: Combining economic analysis with expert elicitations to inform climate policy. Energy Economics **31**, S37–S49 (2009)
6. Baker, E., Chon, H., Keisler, J.M.: Advanced nuclear power: Combining economic analysis with expert elicitations to inform climate policy (2008)
7. Baker, E., Chon H.and Keisler, J.: Carbon capture and storage: combining economic analysis with expert elicitations to inform climate policy. Climatic Change **96**, 379–408 (2009)
8. Baker, E., Keisler, J.: Cellulosic biofuels expert views on prospects for advancement. Energy **36**, 595–605 (2011)
9. Baker, E., Solak, S.: Management of energy technology for sustainability: how to fund energy technology research and development. Prod.Oper.Manag. **23**, 348—365 (2014)
10. Bolger, F., Rowe, G.: The aggregation of expert judgment: Do good things come to those who weight? Risk Analysis (2014)
11. Bosetti, V., Catenacci, M., Fiorese, G., Verdolini, E.: The future prospect of pv and csp solar technologies: An expert elicitation survey. Energy Policy **49**, 308–317 (2012)
12. Chan, G., Anadon, L.D., Chan, M., Lee, A.: Expert elicitation of cost, performance, and rd&d budgets for coal power with ccs. Energy Procedia pp. 2685–2692 (2011)
13. Chung, T., Patio-Echeverri, D., Johnson, T.L.: Expert assessments of retrofitting coal-fired power plants with carbon dioxide capture technologies. Energy Policy **39**, 5609–5620 (2011)
14. Clarke, L., Baker, E.: Workshop report: Rd&d portfolio analysis tools and methodologies. Tech. rep. (2011). URL http://www.globalchange.umd.edu/publications/656/
15. Cooke, R.M.: Messaging climate change uncertainty. Nature Climate Change **5**(8-10) (2015)

16. Cooke, R.M., Goossens, L.H.J.: Tu delft expert judgment data base. special issue on expert judgment. Reliab. Eng. Syst. Safe. **93**, 657—674 (2008)
17. Curtright, A., Morgan, M., Keith, D.: Expert assessment of future photovoltaic technology. Environmental Science and Technology **42**, 9031–9038 (2008)
18. Fiorese, G., Catenacci, M., Bosetti, V., Verdolini, E.: The power of biomass: experts disclose the potential for success of bioenergy technologies. Energy Policy **65**, 94–114 (2014)
19. Fiorese, G., Catenacci, M., Verdolini, E., Bosetti, V.: Advanced biofuels: Future perspectives from an expert elicitation survey. Energy Policy **56**, 293–311 (2013)
20. Gallagher, K.S., Anadon, L.D.: DOE budget authority for energy research, development, & demonstration database. Tech. rep., Energy Technology Innovation Policy research group, Belfer Center for Science and International Affairs, Harvard Kennedy School (2014). URL `http://belfercenter.ksg.harvard.edu/publication/25702/`
21. Goulder, L.H., Schneider Stephen, H.: Induced technological change and the attractiveness of co2 emissions abatement. Resource and Energy Economics **21**, 211–253 (1999)
22. Hall, B.H.: Measuring the returns to r&d: The depreciation problem, nber working paper no. w13473. Tech. rep., Cambridge, MA, National Bureau of Economic Research (2007)
23. Henrion, M., Granger Morgan, M.: Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York (1990)
24. Hoffert, M.I., Caldeira, K., Benford, G., Criswell, D.R., Green, C., Herzog, H., Jain, A.K., Kheshgi, H.S., Lackner, K.S., Lewis John S.and Lightfoot, H.D., Manheimer, W., Mankins, J.C., Mauel, M.E., Perkins, L.J., Schlesinger, M.E., Volk, T., Wigley, T.M.L.: Advanced technology paths to global climate stability: Energy for a greenhouse planet. Science **298**, 981–987 (2002)
25. IPCC: Special report of the intergovernmental panel on climate change on renewable energy sources and climate change mitigation. Tech. rep., CAMBRIDGE UNIVERSITY PRESS (2011)
26. IRENA: Renewable power generation costs in 2014. international renewable energy agency (January 2015)
27. Jaffe, A.B., Newell, R.G., Stavins, R.N.: A tale of two market failures: Technology and environmental policy. Ecological Economics **54**, 164–174 (2005)
28. Jamasb, T.: Technical change theory and learning curves: Patterns of progress in energy technologies. The Energy Journal **28**(3), 51—71 (2007)
29. Jenni, K., Baker, E., Nemet, G.: Expert elicitations of energy penalties for carbon capture. International Journal of Greenhouse Gas Control **12**, 136–145 (2013)
30. Jones, C.I.: R & d-based models of economic growth. Journal of Political Economy **103**, 759–784 (1995)
31. Klaassen, G., Miketa, A., Larsen, K., Sundqvist, T.: The impact of r&d on innovation for wind energy in denmark, germany and the united kingdom. Ecological Economics **54**(2—3), 227—240 (2005)
32. Kriegler Elmar; Blanford, G.J., Weyant John P.; Volker, K., Clarke, L., Edmonds, J., Fawcett, A., Gunnar, L., Riahi, K., Richels, R., StevenK, R., Tavoni, M., Vuuren, D.P.: The role of technology for achieving climate policy objectives: Overview of the emf 27 study on global technology and climate policy strategies. Climatic Change pp. 1–15 (2014)
33. NRC: Prospective evaluation of applied energy research and development at doe (phase two). Tech. rep., NRC (2007)
34. Popp, D., Newell, R.: Where does energy r&d come from? examining crowding out from energy r&d. Energy Economics **34**(4), 980—991 (2012)
35. Qiu, Y., Anadon, L.D.: The price of wind power in china during its expansion: Technology adoption, learning-by-doing, economies of scale, and manufacturing localization. Energy Economics **34**(3), 772—785 (2012)
36. Rao, A., Rubin, E., Keith, D., Morgan, M.: Evaluation of potential cost reductions from improved amine-based co2 capture systems. Energy Policy **34**, 3765–3772 (2006)
37. Rogelj, J., L. McCollum, D., ONeill, B.C., Riahi, K.: 2020 emissions levels required to limit warming to below 2c. Nature Climate Change **3**, 405–12 (2013)
38. Tversky, A., Kahneman, D.: 1974. judgment under uncertainty: Heuristics and biases. Science **185**(4157), 1124—1131. (1974)

39. USEPA: Expert elicitation task force, white paper. Tech. rep., U.S. Environmental Protection Agency (2011)
40. Usher, W., Strachan, N.: An expert elicitation of climate, energy and economic uncertainties. Energy Policy **61**, 811–821 (2013)
41. Verdolini, E., Anadon, L.D., Lu, J., Nemet, G.: The effects of expert selection, elicitation design and r&d assumptions on experts' estimates of the future costs of photovoltaics. Energy Policy **80**, 233—243 (2015). Submitted
42. Zickfeld, K., Morgan, M., Frame, D., Keith, D.: Expert judgments about transient climate response to alternative future trajectories of radiative forcing. Proc Natl Acad Sci **107**(28), 12,451—12,456. (2010)
43. Zubaryeva, A., Thiel, C.: Analyzing potential lead markets for hydrogen fuel cell vehicles in europe: Expert views and spatial perspective. International Journal of Hydrogen Energy **38**, 5878–15,886 (2013)
44. Zubaryeva, A., Thiel, C., Barbone, E., Mercier, A.: Assessing factors for the identification of potential lead markets for electrified vehicles in europe: expert opinion elicitation. Technological Forecasting and Social Change **79**, 1622–1637 (2012)