Climatic Change (2008) 90:189–215
DOI 10.1007/s10584-008-9458-1

Improving the way we think about projecting future energy use and emissions of carbon dioxide

M. Granger Morgan · David W. Keith

Received: 20 March 2007 / Accepted: 4 April 2008 / Published online: 29 August 2008

Abstract A variety of decision makers need projections of future energy demand, CO_2 emissions and similar factors that extend many decades into the future. The past performance of such projections has been systematically overconfident. Analysts have often used scenarios based on detailed story lines that spell out "plausible alternative futures" as a central tool for evaluating uncertainty. No probabilities are typically assigned to such scenarios. We argue that this practice is often ineffective. Rather than expanding people's judgment about the range of uncertainty about the future, scenario-based analysis is more likely to lead to systematic overconfidence, to an underestimate of the range of possible future outcomes. We review relevant findings from the literature on human judgment under uncertainty and discuss their relevance to the task of making probabilistic projections. The more detail that one adds to the story line of a scenario, the more probable it will appear to most people, and the greater the difficulty they likely will have in imagining other, equally or more likely, ways in which the same outcome could be reached. We suggest that scenario based approaches make analysts particularly prone to such cognitive biases, and then outline a strategy by which improved projections, tailored to the needs of specific decision makers, might be developed.

For those of us who work on climate and energy policy it would be extremely useful to be able to predict a few simple things such as the future demand for energy and the future mix of energy technologies over the coming decades—if not as sharp point estimates, then at least as well-calibrated subjective probability distributions. However, the track-record of past efforts to make such predictions is anything but

M. Granger Morgan (✉) · D. W. Keith
Department of Engineering and Public Policy, Carnegie Mellon University,
5000 Forbes Ave, Pittsburgh, PA, 15213, USA
e-mail: granger.morgan@andrew.cmu.edu

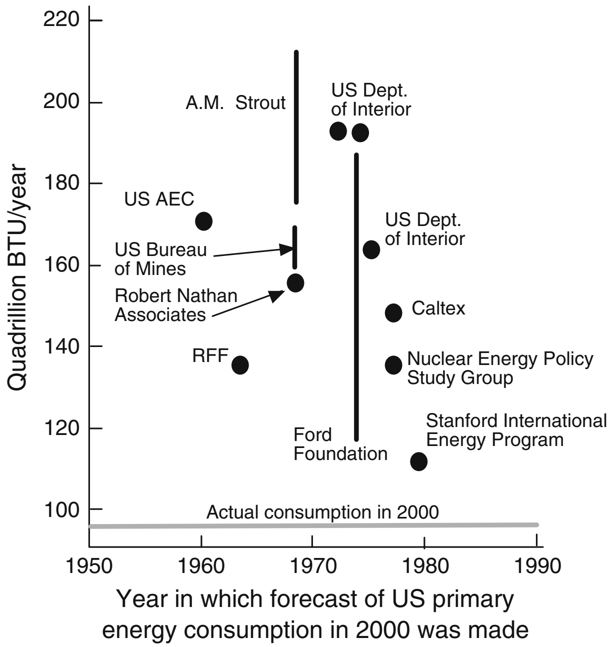
D. W. Keith
Department of Chemical and Petroleum Engineering and Department of Economics,
University of Calgary, Calgary, Canada

reassuring. For example, Smil (2003) has compiled a number of point and interval estimates made by various groups between 1960 and 1980 that projected US primary energy demand in the year 2000. None of these projections included the actual value (Fig. 1). He found similarly poor performance for estimates of global primary energy. Greenberger complied projections made in the 1980s for US energy demand in the year 2000 (Fig. 2). A comparison of these results with actual demand in 2000 yields similarly disappointing performance. An analysis by Craig et al. (2002) reports similar poor results. While this abysmal track record has not stopped researchers and policy analysts from continuing to make projections, or from talking in briefings and papers as though they believe them, it does raise obvious methodological and practical questions. This paper addresses three issues: (1) What do we need to project about future energy technology, energy use, and emissions of greenhouse gasses? What form should such projections take to be useful? (2) What are the problems with the way we have been making such projections in the past? Specifically, why are scenarios with highly detailed "story lines", generally *not* a good way to make such projections? (3) How might we do a better job of making such projections?

1 Why, and for whom, do we need energy projections?

While projections, forecasts or scenarios may be produced by a variety of organizations, and may vary radically in their complexity and content, in our view such projections have no other sensible purpose than to inform current decisions. To be

Fig. 1 Summary of forecasts of US primary energy consumption for the year 2000 compiled by Smil (2003) as a function of the date on which they were made. Figure redrawn from Smil (2003)



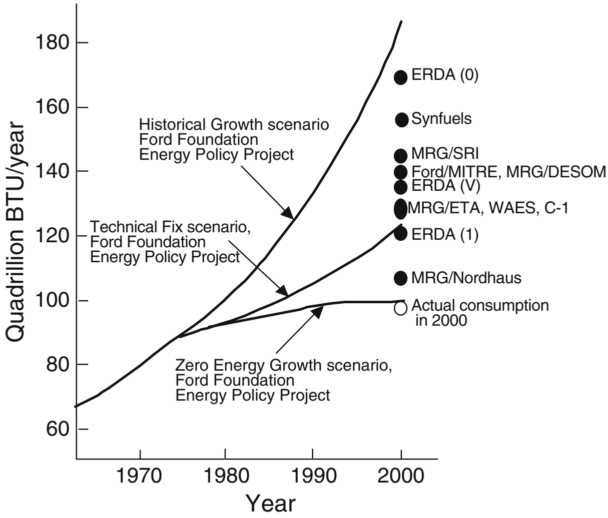


Fig. 2 Comparison of forecasts of US primary energy consumption for the year 2000 compiled by Greenberger in the early 1980s compared with three scenarios developed by the Ford Foundation Energy Project, a major undertaking in the early 1970s. The Ford, and the other studies reported in this diagram, is discussed in some detail in Greenberger (1983) from which this figure is redrawn

effective, projections must therefore be shaped to meet the needs of specific decision makers. Because decision makers and the decisions they face are so heterogeneous, it may be impractical to produce projections which are both generic and useful.

Consider projections of energy futures used in climate policy. Much effort has been centered around the IPCC's Special Report on Emissions Scenarios (SRES) and related analysis (Nakicenovic et al. 2000). The SRES scenarios aim to illustrate the range of emissions in a business as usual case, that is, one with no policies restricting CO_2 emissions. While years ago such scenarios might have been useful for a rough exploration of just how serious the issue of climate change may be, today it is no longer clear what class of decision makers would find such scenarios directly relevant. Following Parson et al. (2006) we distinguish three salient groups making climate related plans and decisions: impacts and adaptation managers, energy resource and technology managers, and finally, government officials and others responsible for developing regional and national policy or national positions in international frameworks.

Impacts and adaptation managers have responsibility for particular climate-sensitive assets, such as land use planning or operation of water supply and irrigation systems, so they need judgments about likely future climate change and its impacts. Whether they represent public or commercial interests, these managers need to incorporate such judgments into their long-range planning. As such, their needs are driven by the expected climate change for which business-as-usual scenarios are inappropriate because such scenarios assume, by definition, that there will be no significant climate policy. Although progress on restraining emissions has been disappointingly slow, the probability that there will be no significant restraint on emissions over the next half-century now seems vanishingly small, particularly given that some restraints already exist. Projections for decision makers who will respond to climate change must take account of the likelihood of future restraints on emissions. In some cases, this may be a dominant uncertainty, in all cases, it cannot be ignored.

Energy resource and technology managers include owners, developers, and investors in energy resources and energy-related capital stock and new technologies. These actors need to explore alternative mitigation policy regimes and their consequences for energy and technology assets on time scales of several decades into the future. In our experience, such predominantly private actors see little direct benefit in century timescale projections of energy technologies or of climate change itself, and are far more concerned with the uncertainty about government regulations and about changes in energy technologies and prices.

Government officials make decisions about policy relevant to both adaptation and mitigation decisions. In the former capacity, they need scenarios of climate change under specified emissions trends, and resultant impacts on regional, national or international communities and resources, particularly key vulnerabilities. In the latter capacity, they need scenarios of aggregate climate impacts, plus scenarios of future emissions, alternative responses and their consequences, and the policy context of mitigation responses in other major nations.

Estimates of emissions on a century timescale under a no climate-policy assumption are useful for understanding the most basic question of climate policy: are climate impacts big enough to worry about if we make no efforts to restrict emissions? However, in our view, that question has been authoritatively answered in the affirmative. Thus, generic century-timescale business-as-usual scenarios have little utility for practical decisions about how to manage the climate problem. As Lave and Epple (1985) argued, in writing about scenario analysis and climate change over twenty years ago "the whole point…is to determine how our actions can lead to a better future. If we cannot affect the future, scenarios might help us to reconcile ourselves to the inevitable, but they would not lead to any change. Thus, it is important to determine the actions and policies that will have the greatest effect on the future and to focus on how those factors might affect it."

2 What's wrong with the way we've been making energy projections?

2.1 The rise of scenario analysis

One reason that energy projections such as those shown in Figs. 1 and 2 have done so poorly may be that forecasters have not thought broadly enough about how the future may be different from the past. The realization that methods are needed to stretch one's thinking about how the future may unfold has been a cited as a central rationale for the development of "scenario analysis," and the construction of scenarios about the future has become a central tool for making projections about future CO_2 emissions for use in climate policy.

For example, for over 30 years the Shell Group (2005) has been developing "global scenarios" in order to "cast light" on the global environment in which the firm may find itself operating and "to identify emerging challenges and to foster adaptability to change." Shell explains that these "…scenarios are used to help review and assess strategy."

These explanations are consistent with the view that projections of the future are tools to aid current decision making. Shell goes on to argue that scenarios are not projections:

> They are not forecasts, projections or predictions of what is to come. Nor are they preferred views of the future. Rather, they are plausible alternative futures: they provide reasonable and consistent answers to the 'what if?' questions relevant to business. (Shell Group 2005)

Some years ago Shell built scenarios around a simple dichotomy. In its latest instantiation, Shell scenarios are designed to help explore the tradeoffs and tensions between a "tricotomy" of "efficiency—market incentives," "social cohesion and justice—the force of community" and "security—coercion, regulation" in a globalized world. Using an attractive triangular graphic they elaborate and explore the implications of the three "possible futures" outlined in Table 1. Then, to explore the implications of these future worlds to the business environment in which they may find themselves, they link the efficiency dimension to "fast energy intensive global growth," the social cohesion and justice dimension to "carbon emissions monitored and priced," and the security dimension to "energy security and concerns over long-term energy supplies and accessibility," and explore how these goals might be achieved in the worlds summarized in Table 1.

Inspired in part by the work of Shell and others, in 1996 The Intergovernmental Panel on Climate Change (IPCC) commissioned a Special Report on Emission Scenarios (widely termed the SRES scenarios; Nakicenovic et al. 2000). The SRES team explained that the scenarios they developed:

> ...cover a wide range of the main driving forces of future emissions, from demographic to technological and economic developments. As required by the Terms of Reference, none of the scenarios in the set includes any future policies that explicitly address climate change, although all scenarios necessarily encompass various policies of other types. The set of SRES emissions scenarios is based on an extensive assessment of the literature, six alternative modeling approaches, and an "open process" that solicited wide participation and feedback from many groups and individuals. The SRES scenarios include the range of emissions of all relevant species of greenhouse gases (GHGs) and sulfur and their driving forces...scenarios are used to help review and assess strategy.

The SRES scenarios were built around "story lines" that involved such details as "economic and cultural convergence and capacity building, with a substantial reduction in regional differences in per capita income" (the A1 scenarios); a world

Table 1 Summary of the three "possible futures" in the year 2025 explored in the latest instantiation of Shell's global scenario efforts

Scenario name	Location in the three dimensional space
Low trust legalistic world (a legalistic "prove it to me" world)	Low on scale of social cohesion and justice—the force of community
Flags (a dogmatic "follow-me" world)	Low on scale of efficiency—market incentives
Open doors world (a pragmatic "know me" world)	Low on scale of security—coercion, regulation

in which "underlying these is self reliance and preservation of local identity" (the A2 scenarios); a world with "rapid change in economic structures toward a service and information economy... [with emphasis on] global solutions to economic, social and

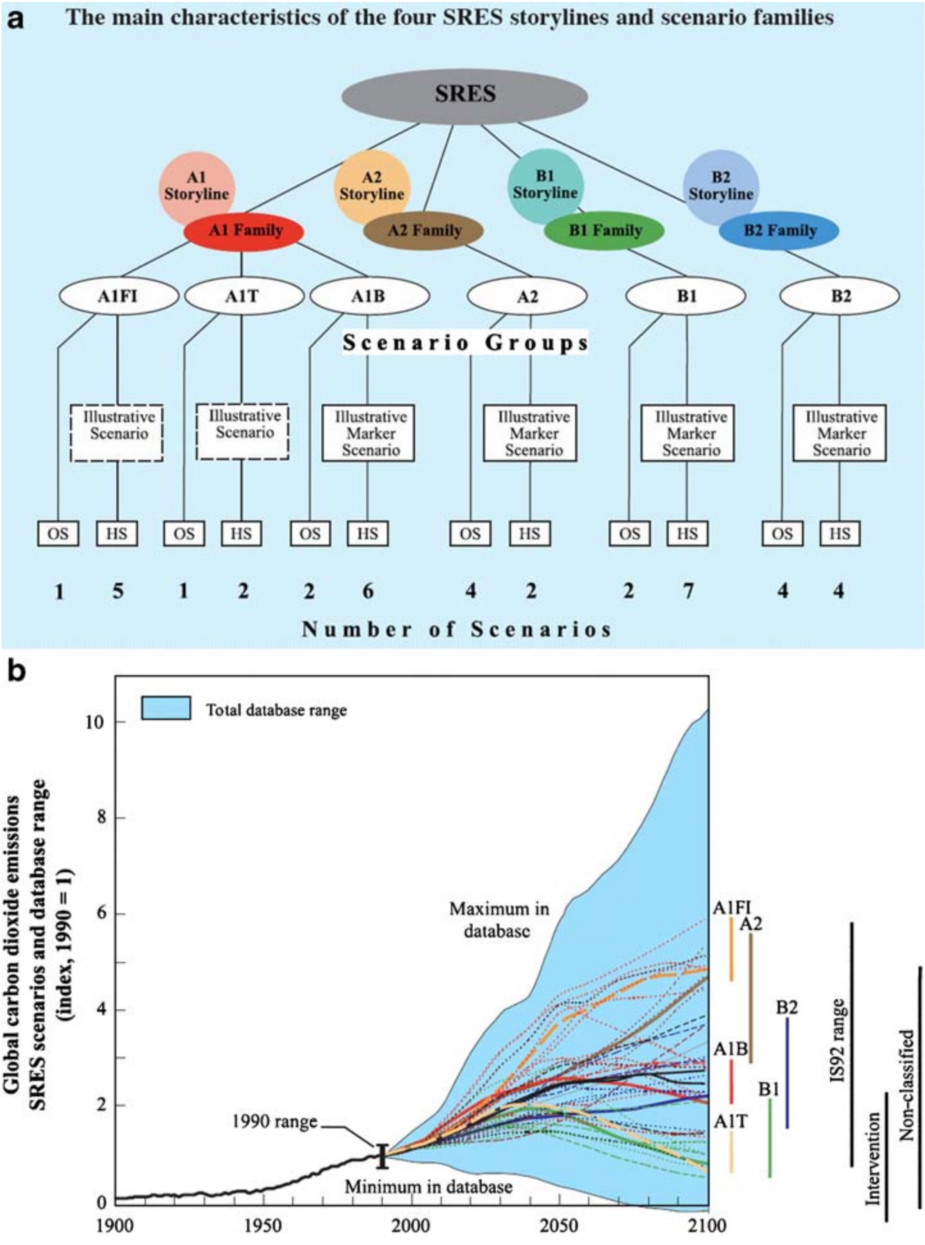


Fig. 3 **a** Basic structure of the SRES scenarios; **b** the resulting projections of CO_2 emissions. Details associated with each story line go on for many pages. A plot similar to **b** is also available for emissions resulting from changes in land use. Reproduced from Nakicenovic et al. (2000)

environmental sustainability" (the B1 scenarios); or a world in which "the emphasis is on local solutions to economic, social and environmental sustainability..." (the B2 scenarios). Associated with each are then projections of future CO_2 emissions as shown in Fig. 3.

While the argument is made by the developers of the SRES scenarios that they are designed to expand one's thinking about the future, the reality is that these scenarios have seen relatively little use in applications other than to use the associated time series projections of CO_2 emissions as inputs to GCMs which are then run to estimate the future evolution of the climate. One exception is work by Arnell et al. (2004) which does adapt and apply the SRES scenarios to a number of specific impact assessments in the UK. However, even in this case, it does not appear that very much of the complicated detail that is provided in the SRES story lines is used or has significant consequences for the resulting analysis.

While the SRES scenarios were developed to display a range of possible energy and emissions futures in the absence of climate policy, it is easy to imagine that in a future round of analysis, IPCC or others may develop a similar set of scenarios that include possible climate policy responses as part of their story lines. Thus, in what follows, while we discuss a number of the limitations of the SRES approach, we use the current scenarios as qualitatively representative of similar scenarios that might be produced in the future with various climate policies included as part of the story lines.

Like Shell, the SRES team argued that their scenarios are not forecasts, saying:

> They are not forecasts, projections or predictions of what is to come. Nor are they preferred views of the future. Rather, they are plausible alternative futures: they provide reasonable and consistent answers to the 'what if?' questions relevant to business (Nakicenovic et al. 2000).

The developers of the SRES scenarios describe each as "equally sound" but argue that no probability should be attached to any given scenario; that they are simply provided as a means of exploring a range of future possibilities.[1] Grübler and Nakicenovic (2001) support this position, combining a frequentist argument together with an argument of infeasibility if one does accept a subjectivist approach:

> Probability in the natural sciences is a statistical approach relying on repeated experiments and frequencies of measured outcomes, in which the system to be analyzed can be viewed as a 'black box'. Scenarios describing possible future developments in society, economy, technology, policy and so on, are radically different.
>
> First, there are no independent observations and no repeated experiments: the future is unknown, and each future is 'path-dependent': that is, it results from a large series of conditionalities ('what if...then' assumptions) that need to be followed through in constructing internally consistent scenarios. Socio-economic variables and their alternative future development paths cannot be combined at will and are not freely interchangeable because of their interdependencies. One should not, for example, create a scenario combining

[1]Recent work by Schweizer and Kriegler (2007), using cross-impact analysis, suggests that some of the SRES scenarios are significantly more internally consistent than others.

> low fertility with high infant mortality, or zero economic growth with rapid technological change and productivity growth—since these do not tend to go together in real life any more than they do in demographic or economic theory. Second, the 'what if…then' approach requires the explicit representation of a system to describe how the variables interact. That is, any randomly assumed distribution of 'what if' conditions (based on expert opinion about the future of world population, economic growth, technological development, diets and so on) would be insufficient without also assigning distributions to the conditional probabilities by which these assumed future states of driving forces interact to influence an unknown outcome (greenhouse-gas emissions). This means that the distributions of 'what if' as well as 'then' remain unknown. (Grübler and Nakicenovic 2001)

While the authors of scenarios may decline to provide explicit statements about probability, in our view, judgments about the subjective probability of future states of the world lies at the heart of scenario construction. The literature on scenarios often aims to make a sharp distinction between scenarios and forecasts or projections; for example, it is asserted that scenarios are judged by their 'feasibility' or 'plausibility' rather than their likelihood. We cannot find any sensible interpretation of these terms other than as synonyms for relative subjective probability. Absent a supernatural ability to foresee the future, what could be meant by a statement that one scenario is feasible and another infeasible but that the first is (subjectively) more probable that the second?

While the presentation of scenarios may differ in style from that of projections or forecasts, all are expressions of the authors' subjective judgments about the relative probability of various futures. In part, the term scenario may be used simply to denote a relatively low degree of confidence in one's ability to make predictions. Parson et al. (2006) note that "although different authors' usage is not consistent, 'prediction' and 'forecast' usually denote statements for which the highest confidence is claimed. 'Projection' denotes a less confident statement, which may have some specified confidence level and may be explicitly contingent on specified assumptions about other future conditions. Calling a future statement a 'scenario' usually implies still less confidence and more associated contingencies."

2.2 When and how can one attach probabilities to scenarios?

When scenarios are useful to decision makers, it is hard to avoid the conclusion that they are useful precisely because they communicate, in some measure, the analyst's judgment about the relative probability of various futures to decision makers.[2] We judge that the frequent statements to the contrary made by analysts arise from (1) discomfort with the necessarily subjective nature of the probability assessment and (2) the need to avoid disputes about the probability judgments that necessarily underlie the choice of scenarios.

[2] While Shell is steadfast in insisting that probabilities should never be attached to the global scenarios that it develops, Albert Bressand, who headed the team that developed the most recent Shell Trilemma Triangle map, has indicated to us in a private communication that in other less global applications outside of the construction of global scenarios he has attached probabilities "not to any point on the map but to specific trajectories" through that triangular space.

Schneider (2001) and others have argued that without probabilities scenarios are of little value to climate scientists and impact assessors who are trying to understand how the climate is likely to evolve over the coming centuries. Schneider suggests that no probabilities were attached to the SRES scenarios simply to "avoid endless disputes" because of the "divergent views of participants [in the scenario development process] of the relative likelihoods." As a participant in the early formulation of these scenarios, he notes that:

> While acknowledging the logic of avoiding fruitless debate, I strongly argued at the time that policy analysts needed probability estimates to assess the seriousness of the implied impacts; otherwise they would be left to work out the implicit probability assignments for themselves. . . . I urged the expert group to provide a subjective probability assessment for less expert users, but I was not persuasive enough, and the SRES authors expressed 'no preference' for each scenario. (Schneider 2001)

Users of the SRES scenarios have sometimes ignored the argument that probabilities should not be attached to the SRES scenarios, and have assigned equal probability to all of them, using them as input for probabilisitic analysis (e.g., Wigley and Raper 2001). This practice is problematic. Analysts who create scenarios do so based on their subjective judgments about probabilities. Users must ascribe some measure of likelihood to the scenarios in order to use them in further analysis or decision-making. If judgments about likelihood are not supplied with the scenarios, they will be assumed by the users either explicitly or implicitly. The convention of not communicating information about the relative likelihood of scenarios therefore muddies communication between analysts and users.

If we think of a scenario as describing a series of points over time through a multidimensional space of future possible socioeconomic conditions, then the developers of most of these scenarios are correct (but for the wrong reasons) that no probability should be assigned to scenarios. Viewed this way, scenarios cannot be assigned probabilities since, in any probability distribution over a continuous variable, the probability that attaches to any specific point value is zero. However, statements about the probability of scenarios *can* be made if the scenarios are specified using *ranges* of values for the socioeconomic variables of interest. The probability that global energy use in 2100 will be *precisely* 1,000 EJ/year must be zero; but it is perfectly sensible to assert under some set of assumptions, that in 2100 there is a 20% chance that the value of global energy use will fall between 800 and 1,200 EJ/year.

Someone who makes such a statement adopts a subjectivist or personalist view of probability,[3] that is that probability is a statement of their degree of belief (de Finetti 1974). In this view, the probability that the person assigns to the occurrence of an event X is defined as $p(X|e)$, where e is all the evidence that person has that is relevant to making the specific judgment. Of course, if e consists of a very large number of instantiations of the exact same random process that will give rise to outcome X, then in that case it would make sense to equate the subjectivist judgment

[3] A subjectivist perspective is often also termed Bayesian. For many readers that term immediately implies that one is updating a prior distribution through the use of Bayes' Rule. That need not be the case. To avoid confusion, we have not used the term Bayesian in this paper.

with the classical or “frequentist” definition of probability. But, when the weather bureau forecasts a 30% ($p = 0.3$) chance of rain, or when someone gives you 3:1 odds ($p = 0.75$) that Italy will beat England in the next world cup they are using subjective probability. The prediction may be based on considerable data, knowledge and past experience, but such probabilities are statements of belief.

It is easy enough to produce scenarios that span a range of values. For example, suppose that one thinks that the value of a quantity of interest called Q, will grow in some monotonically increasing way over time, as shown in Fig. 4a. Then, if we can assess a probability distribution for that value at one or a number of moments in the future, we could discretize the results into a number of associated scenarios, as shown in Fig. 4b, in which each scenario represents some interval across the range of future possible values of Q. Similarly, if one entertains more than one model of the way in which $Q(t)$ could evolve, that too can be handled with a number of discretized partitionings of the space. A simple example of this is shown in Fig. 5.

2.3 Simpler ways of thinking about alternative futures

It is possible to use a self-consistent treatment of subjective probabilities in scenario analysis without explicitly producing probability distributions. In the US National

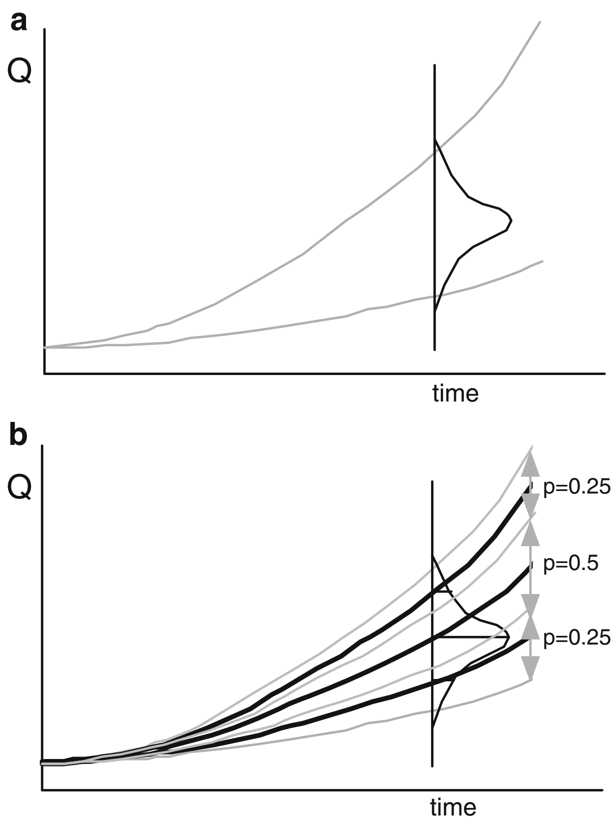


Fig. 4 **a** The future uncertain evolution of a quantity Q can be discretized and represented (**b**) by a family of several curves, to each of which is associated a specific probability which together sum to 1. For simplicity of illustration, here we show just three such curves. Of course, if greater resolution is needed, for example, because the result will be used as input to a very non-linear model, more discretized curves could be used

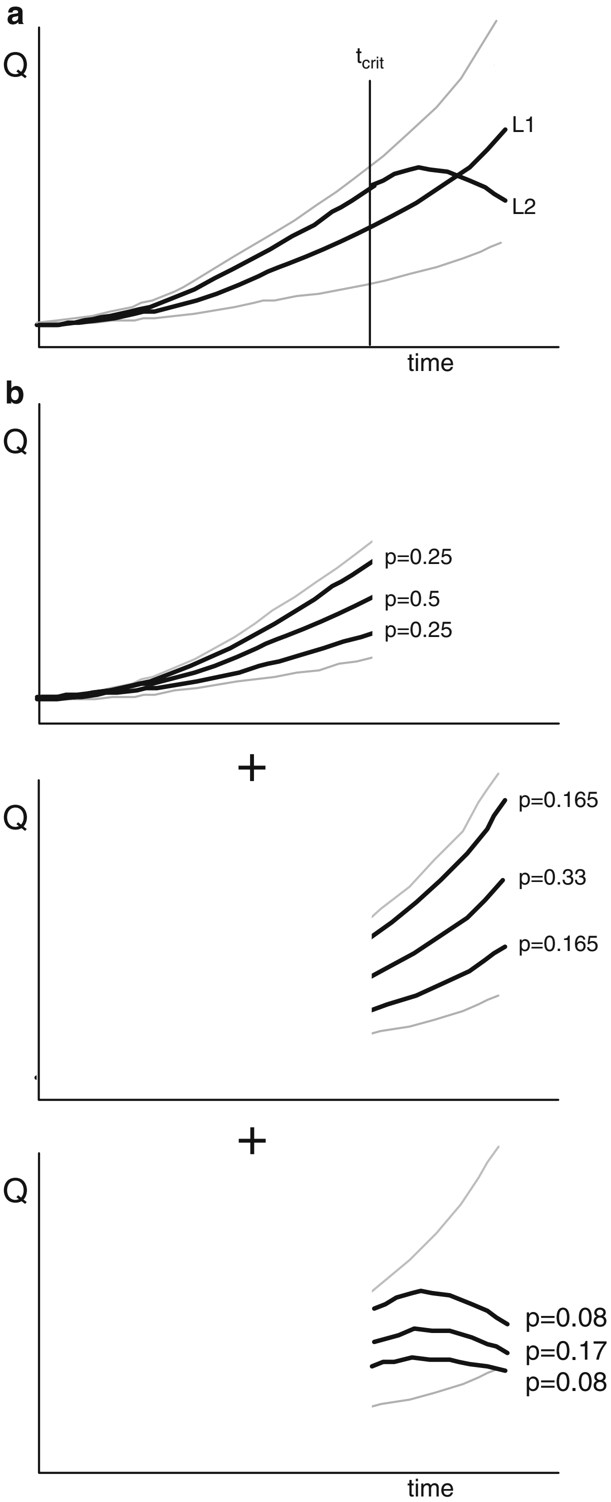


Fig. 5 **a** In contrast to Fig. 4, suppose that after some future time t_{crit}, there is a possibility that the quantity Q may evolve in two very different ways, here represented by curves $L1$ and $L2$. **b** While the results become a bit more complex, the same strategy can be used to decompose the result into a number of discrete curves, each of which represents a range of value in $Q(t)$ and to each of which is associated a specific probability

Assessment, a simple set of basic projections to the year 2050 of population, income and employment was developed. Then each regional and sectoral team was asked to select "*a few key issues* that they judged would be most important for the region or sector" (US National Assessment 2001). For each key issue:

> …the team was asked to identify *one or two key socioeconomic factors*, such as specific aspects of development patterns, land use, technologies, or market conditions, that they judged likely to have the most direct influence on climate impacts, capacity for adoption and vulnerability for that issue...[teams] were then to examine the impacts of specified climate-change scenarios on their key issues, under a range of values for their chosen socioeconomic factors. If they identified more than one key socioeconomic factor, they were asked to construct a few alternative socioeconomic scenarios by varying the factors jointly across high and low values...
> …teams were asked to make the range wide enough to generate instructive variation in impacts, but to remain within their judgment of plausibility. Specifically, the range chosen for any factor should correspond to roughly a 10% chance that the true value would lie above the upper end of the range, and a 10% chance that it would lie below the lower end. (US National Assessment 2001)

In this example, participants were asked to choose ranges for the factors they identified using a well-posed statement about subjective probabilities (the 80% confidence interval). In scenario analysis, individual participants or modeling teams are continually engaged in choices about the range of parameter variation to be included in the analysis. If the choice of parameter range is made without some agreed definition of the likelihood which the range is designed to encompass, then it is hard to meaningfully interpret the results.

Of course, specifying a meaningful 80% confidence interval requires one to think systematically about *all* the ways that given values might arise. As outlined in the next section, the literature in experimental psychology suggests that doing this is a difficult cognitive task.

3 Human judgment under uncertainty

There is a large literature in experimental psychology demonstrating persistent overconfidence when lay people and experts make judgments in the face uncertainty. Thus, psychologists would find nothing surprising about the poor performance of the energy forecasts summarized in Figs. 1 and 2. A standard way to report overconfidence is with the "surprise index," the fraction of true values that lie outside respondents' 98% confidence intervals when people give probabilistic estimates of quantities whose values can be known. Figure 6 shows a compilation of results from 21 studies involving thousands of subjects in studies run by 17 investigators, all of which display overconfidence (Lichtenstein et al. 1982; Morgan and Henrion 1990). Unfortunately, in many contexts, experts are subject to the same systematic bias (Slovic et al. 1980). For example, Fig. 7 shows best-recommended values for the speed of light made since the 1930s, a number of which do not include the best modern value. Similar results have also been found for a number of other fundamental physical constants (Henrion and Fischhoff 1986).

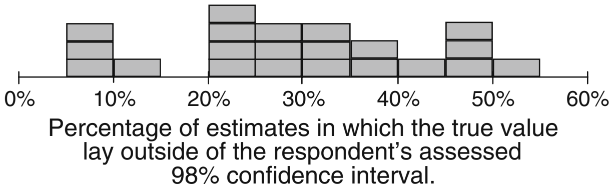


Fig. 6 Illustration of overconfidence in probabilistic judgments. In 21 separate studies, well-educated people were asked to make judgments about the value of a large number of known quantities (such as the length of the Panama canal). They were also asked to provide a 98% confidence interval on those judgments. The histogram reports the proportion of the time that the true answers lay outside the 98% confidence interval the respondents had given, which, of course, should have been 2% (each box in the histogram reports the results of a separate study, several of which had more than 1,000 participants). Histogram constructed from data summarized in Morgan and Henrion (1990)

One reason for overconfidence is that people find it very difficult to recall factors or situations that are not readily "available" for recall. For example, Fischhoff et al. (1978) ran an experiment in which they presented respondents with a fault tree illustrating reasons why a car would not start. They constructed several different versions of the tree, moving different portions of the tree into a branch labeled "all other problems" and asked respondents to judge the completeness of the representation. They found that when they placed varying portions of the tree into the category "all other problems" the probability that respondents attached to that

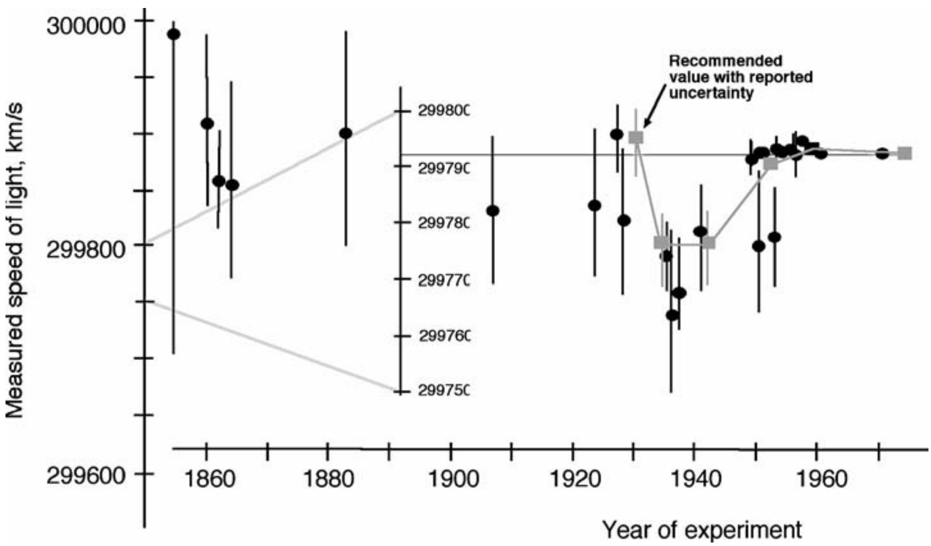


Fig. 7 History of reported experimental values of the speed of light (*black dots*) and "recommended values" (*gray boxes*) with associated uncertainties. The uncertainty ranges associated with the recommended values are an indication of the subjective probability distributions of the expert community. The fact that a number of these ranges of recommended values do not include the actual value illustrates the fact that lay people are not the only ones subject to overconfidence. Figure constructed from data in Henrion and Fischhoff (1986) who report similar results for a number of other physical constants

catch-all category was far less sensitive to what had been left out than it should have been. For example, in a typical case, the probability only doubles when it should have increased by a factor of six. They also found that as a category was elaborated people assigned greater probability, perhaps suggesting “that people tend to assign some minimum probability in any category with which they are faced” (Fischhoff et al. 1978).

Fischhoff et al. (1978) also administered their survey to a group of professional automobile mechanics. While the effect in this case was weaker, even this group showed too little sensitivity to how much of the trimmed tree was included in the category “all other problems,” under estimating the probability by roughly a factor of 2. “Neither self-rated degree of knowledge nor actual experience had any systematic relation to ability to detect what was missing from the tree”.

Tversky and Koehler (1994) have generalized these findings in a large group of studies in which, in the context of a theoretical framework they term “support theory,” they show that the amount of probability assigned to contingencies that are not explicitly listed is almost always too low, and decreases somewhat as more detail is provided about those categories that are presented (Fig. 8).

For both expert and lay respondents the availability heuristic[4] appears to dictate that “what is out of sight is effectively out of mind” (Fischhoff et al. 1978; Slovic et al. 1980) suggesting the need to explicitly identify and list factors when people are asked to make such judgments. Recently, Fox and his co-authors have advanced experimental evidence that in making judgments across a specified set of categories respondents may start with an “ignorance prior,” assigning equal probability to each articulated branch or option, and then, in accordance with the heuristic of anchoring and adjustment,[5] fail to sufficiently adjust their probabilities in light of specific knowledge they have about the options (Fox and Clemen 2005; Fox and Rottenstreich 2003).

We turn now to a set of findings that are even more troubling for SRES-style scenario analysis. There is a substantial body of evidence that the availability heuristic can lead people to overestimate the probability of a scenario or story line when the detail with which it is specified is increased. This may have serious adverse implications for the use of scenarios for decision support.

[4]The “availability heuristic” is one of several well documented cognitive heuristics (Kahneman et al. 1982) that humans unconsciously use when making judgments about uncertainty. In this case, people assess probabilities of a future event or outcome on the basis of how easily they can remember past examples or how easily they can imagine possible examples. Thus, for example, people tend to overestimate the risk of death from very rare events that are widely reported in the press (e.g. botulism poisoning) and underestimate the risk of dying from very common events which are reported much less frequently than they actually occur (e.g. stroke; Lichtenstein et al. 1978).

[5]“Anchoring and adjustment” is another well-documented cognitive heuristics (Kahneman et al. 1982). In making judgments about uncertain events, people often start with an initial value or “anchor” and then modify their judgment as they consider factors relevant to the specifics of the issue at hand. Often this adjustment is insufficient, even when the respondent knows that the initial anchor was arbitrary (Kahneman et al. 1982). Mullin (1986) has shown that when making judgments in domains in which they are very knowledgeable, experts may sometimes use anchoring and adjustment as an effective strategy to improve the quality of their estimates.

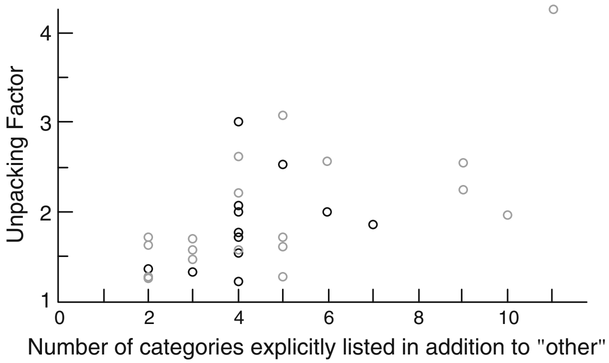


Fig. 8 Summary of "unpacking factor" plotted from data reported by Tversky and Koehler (1994) from 11 studies reported in five different papers, in which subjects were asked to judge the probability of explicitly listed qualitative factors such as causes of death plus a category "all other" (*black circles*) and 19 studies reported in six different papers in which subjects were asked to make judgments of quantitative factors such as probability that a randomly selected adult is between 6′ and 6′ 2″ tall (*gray circles*). Tversky and Koehler (1994) define the "unpacking factor" as the ratio of the sum of all the probabilities elicited by subjects given the enumerated set of categories plus "all other" to the probability judgment by subjects not given an elaborated list. They note that "subadditivity in the judgments is indicated by an unpacking factor greater than 1...." Whether the questions were posed in terms of probability or odds reportedly made little difference. On the other hand, when subjects are given an exhaustive list, elicited probabilities sum essentially to 1

The problem is clearly illustrated in a study conducted by Slovic et al. (1976) in which subjects were given the following personality sketch of a student named Tom W:

> Tom is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others.

In a first treatment, subjects were asked to assess the probability that "Tom W. will select journalism as his college major." In a second treatment, subjects were asked what is the probability that "Tom W. will select journalism as his college major but become unhappy with his choice?' Finally, in a third treatment, subjects were asked what is the probability that "Tom W. will select journalism as his college major but become unhappy with his choice and switch to engineering?" Since each of the latter cases is a *subset* of the earlier cases, the assessed probabilities should *decline* as more and more details are added. In fact, the reverse was found, The assessed probabilities *grew* from 0.21 for the simplest case to 0.39 and then to 0.40 for the most detailed "story line."

Lest one conclude that this dismaying result is the consequence of some unique detail of the study design, Tversky and Kahneman (1983) ran a whole series of studies in which they tried various manipulations designed to induce respondents to give responses that did not reflect such a "conjunction fallacy." They began with some “indirect and subtle” questions in which subjects were asked to judge the probability of simple and compound events. They:

> …expected that even naive respondents would notice the repetition of some attributes, alone and in conjunction with others, and that they would apply the conjunction rule and rank the conjunction below its constituents. This expectation was violated, not only by statistically naïve undergraduates, but even by highly sophisticated respondents. (Tversky and Kahneman 1983)

They then ran a series of studies that made the conjunction more and more obvious. In the simplest formulation, they provided a personality sketch of a woman named Linda and then asked respondents to indicate which of the following two statements was more likely to be true:

> “Linda is a bank teller.
> Linda is a bank teller and is active in the feminist movement.”

Eighty-five percent of respondents chose the latter statement as more probable!

Subsequent manipulations designed to explain and focus respondent’s attention on the “conjunction fallacy” never managed to reduce the fraction of wrong answers below 56%. Tversky and Kahneman (1983) have demonstrated similar problems among several expert groups, such as physicians engaged in making medical judgments about the relationship between symptoms and disease. Redelmeier et al. (1995) extended that work with physicians, to demonstrate similar findings in much greater detail. Fox and Birke (2002) have conducted detailed studies in which they asked lawyers to make judgments about trial outcomes, and again obtained similar results.

Tversky and Kahneman (1983) have demonstrated similar findings in studies involving scenarios about the future. For example, subjects found the conjunction of an earthquake and a flood in California in which 1,000 people drown some time in 1983 roughly 40% more probable than the simpler case of a flood somewhere in North America in which 1000 people drown some time in 1983.

Finally, and perhaps most relevant to our concerns, are the studies Tversky and Kahneman (1983) conducted with 115 participants at the July 1982 Second International Congress on Forecasting in which they repeatedly found violations of the conjunction rule in probabilistic judgments involving comparisons between simple and compound forecasts. For example, the probability of “a 30% drop in the consumption of oil in the USA in 1983” was judged to be 0.22 while the probability of “a dramatic increase in oil prices and a 30% drop in the consumption of oil in the USA in 1983” was judged to be 0.36.

Gregory (2001) has argued that:

> Practitioners can find several advantages in using scenarios. First, they can use scenarios to enhance a person’s or group’s expectancies that an event will occur. This can be useful for gaining acceptance of a forecast….Second, scenarios can be used as a means of decreasing existing expectancies….Third…scenarios can produce greater commitment in the clients to taking actions described in them.

Gregory supports these claims by citing much of the same literature that we have described above. Surreptitiously shaping someone’s views through such manipulation may be standard practice in political or advertising settings. In our view, it is *not* appropriate in policy analysis, where the objective should be to give analysts, planners and decision makers balanced and unbiased assessments on which to base their decisions. We read Gregory's enthusiastic endorsement of the advantages of using one or a few detailed scenarios as a compelling argument against their use in policy analysis!

The largely separate literature on war gaming also suggests problems that may arise from the use of detailed scenarios (Goldberg et al. 1990; Brewer and Shubik 1979; Bracken 1977). There has long been a concern that participants in such game can over generalize from the specific story line developed and played in the game. Bracken (1977) performed a most interesting analysis of ways in which the use of scenarios in strategic war gaming has led to unintended and sometimes undesirable consequences. He analyzed a series of military war gaming examples and identified examples of what he terms "diverting, learning, and suppression." By diverting he means, a process by which war gaming, and the often-detailed scenarios on which such games are based, may divert attention from key elements that deserve attention. Learning refers not just to the learning that specific participants take from the experience but also to a collective view that may emerge, often quite indirectly and unconsciously, among decision makers. While Bracken identifies some instances in which this has been a very positive development, he also identifies a number of cases in which the consequences were extremely negative, noting, in a more recent paper that "gaming and simulation can also reinforce biases and narrow the span of an organization's attention" (Bracken 1990). Finally, on suppression Braken notes:

> There is a well-known tendency for individuals to suppress both unpleasant memories and future possibilities. . . .The extension from individual suppression to group and, in some instances, even organizational suppression is a surprisingly easy step. . . .
>
> It is convenient to consider two forms of suppression. . . outcomes that for some reason are undesirable may be suppressed and this can be termed suppression by commission. Alternatively, suppression may occur by omission. The omission of problem, scenarios or strategies. . . can be a convenient method for avoiding unpleasant eventualitiesm. (Bracken 1977)

There is clearly a risk of suppression occurring when scenarios for use in energy and climate analysis are constructed in a public setting in which future outcomes that include developments such as negative economic development in some regions, pandemics, or regional nuclear war, are excluded because they are politically unacceptable.

4 How might we do a better job of performing future energy and emissions projections?

In Section 2, we argued that projections or scenarios should be developed in such a way as to gives users some idea of the associated probability of each. That means that they should be defined to span a region in the space of interest, and not just a single point in, or line through, that space. From the literature reviewed in Section 3, we conclude that the more detail that one adds to the story line of a scenario, the more probable it will appear to most people, and the greater the difficulty they are likely to have in imagining other ways in which variables of interest might arrive at the same values.

From these insights we conclude that if one wants to assess a probability distribution on the future value of a quantity Q (such as energy consumption or CO_2 emissions) that depends on a variety of complex and uncertain factors, one should *not* start by developing a few complex story lines about the future. Rather, one should employ a strategy that starts by listing as many different ways in which Q might have a high, medium or low value by the future time(s) of interest. Since, as we saw in Section 2, "out of sight is often out of mind" this process should involve multiple participants and be iterative. For example, having carefully specified the quantity and time(s) of interest, one might ask several different expert analysts to each independently take a few days to think about and fill in such a matrix. Then one might put them all together to brainstorm, after which the procedure could be repeated one or more additional times. The objective in this first phase of the effort would *not* be to assess probabilities, but simply to build as full a set as possible of developments and factors that could influence the future value of Q.

4.1 Decomposing the judgments

The most effective strategy for estimating the value of a complex quantity is to decompose the task into various simpler components each of which can be evaluated independently and assigned a value or probabilistic range of values. For example, one might estimate the energy use in 2050, by decomposing the task into various subtasks using the Kaya Identity and estimating the population, GDP per capita and energy use per unit GDP.

The use of models may be viewed as an extension of the decomposition strategy, though dynamic models go beyond decomposition. Their use is discussed in Section 4.3 below.

Within the decision analysis community decomposition has long been a standard strategy for improving probabilistic judgments, although formal experimental studies have yielded mixed results as to its effectiveness (Morgan and Henrion 1990; MacGregor 2001). Decomposition is also a standard strategy employed in order-of-magnitude reasoning of the sort often employed by natural scientists and engineers (Harte 1985; Mullin 1986).

In one effort to explore the advantages of decomposition Armstrong et al. (1975) conducted a study in which half of their subjects were asked to assess the value of a variety of summary quantities, such as "How many packs of Polaroid color film were used in the USA in 1970?" while half were asked a series of component questions such as US population in 1970, persons per family, percentage owning cameras, etc. from which the authors then computed the answer. Armstrong (1978) reports that "the decomposed version yielded more accurate forecasts for all questions, and the improvements were much greater for questions in which uncertainty was high...." MacGregor and Armstrong (1994) have also conducted studies of the benefits of disaggregation and concluded that decomposition improved accuracy when the value being estimated was very large or very small, and involved considerable uncertainty.

4.2 Considering multiple possibilities and multiple future paths

In addition to arguing for the use of decomposition, Armstrong (1978) argues that forecasters should adopt an approach of "eclectic research," in which they should employ a number of different strategies to gain insight about the same issue. He notes, "eclectic research can be applied in a variety of ways in forecasting. It can be used to measure variables in a forecasting model... or relationships in a forecasting model... or it can be used to assess the validity of a relationship... or the uncertainty involved with a forecast" (Armstrong 1978).

In doing this for something as complex as the future of the energy system, it would also be important to think systematically about possible path dependencies. Much of the current structure of the energy system, and of other large-scale societal system architectures is not the result of a single choice or decision process. For example, in the 19th century, nobody decided that all cities should have sanitary sewers. Rather, individuals installed running water before sewer systems were common. This created a major problem that few had thought about or anticipated, precipitating the need for municipalities to scramble to install sewer systems (Tarr 1996). In the twentieth century, nobody decided that natural gas infrastructure should be developed to support home heating in cities such as Pittsburgh. Rather, this was an emergent consequence of the development of town gas networks and construction of the "big inch" and "little inch" oil and gasoline pipelines, that ran from East Texas and the Gulf States to the North East in order to avoid submarine attacks on tankers and assure fuel supplies during the Second World War. The post-war conversion of those lines to natural gas in the face of concerns about international oil prices, and the need to find a home heating fuel that was cleaner than soft coal, was undertaken in order to address an air pollution problem that had grown to critical proportions (Yergin 1991; Tarr 1996). Much of our infrastructure, and many of today's social systems, have emerged in similar ways.

Suppose that one is concerned about estimating various quantities (such as CO_2 emissions) related to the US electric power system some time after the mid-point of the current century. To do that one would need to make judgments about the technical mix that will supply electric power. While a "wedge" (Pacala and Socolow 2004) or similar model (Edmonds et al. 2000; EPRI 2007) may be a conceptually attractive way to think about energy supply over the next 50 to 100 years, history suggests that a model of sequential bulges, in which first one and than another energy technology dominates, may be more realistic.

Three factors that clearly could influence how the mix of technologies in the power system evolves over the coming 50+ years are: the price of natural gas as compared with coal or other fuels or renewable resources; the timing and stringency of future regulatory constraints on CO_2 emissions; and the level of concern that develops about supply reliability and security.

While there are certainly other factors, including public attitudes, that could also play important roles, these three are sufficient to illustrate the basic idea. First, one could identify a number of changes in the external world, and in available electric power technology, that could result in significant impacts on each of the factors of interest. Then, one could systematically consider the consequences of these changes occurring in a variety of simple time sequences. Figure 9 illustrates just four of the possible trajectories. Note that some combinations might lead to the same end-point

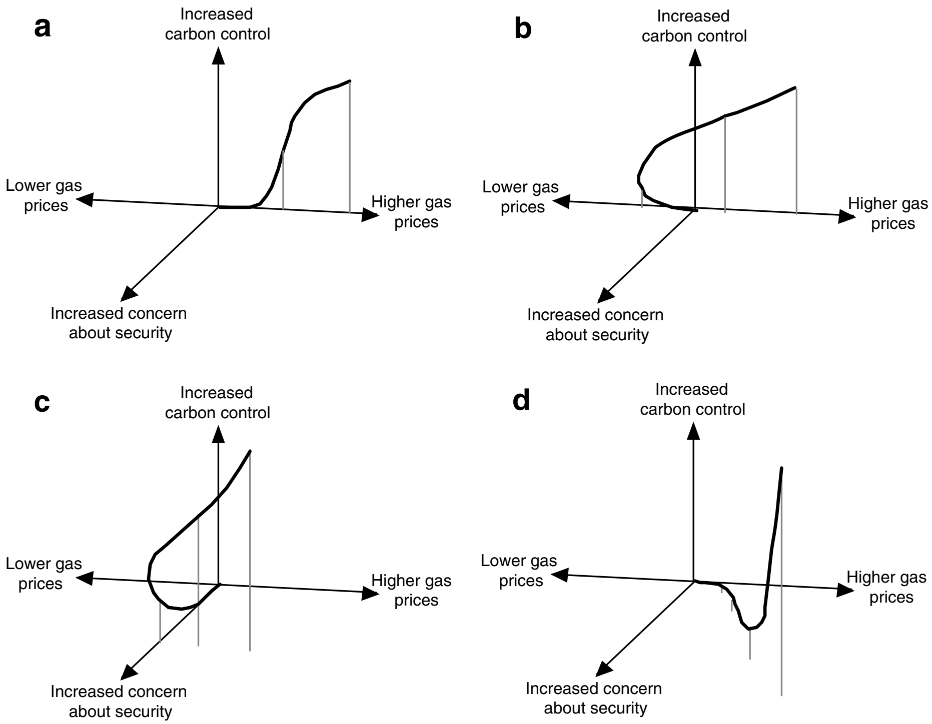


Fig. 9 Four possible future time trajectories through a space of three factors likely to shape the evolution of the electricity system. **a** If gas prices first get high and then carbon control becomes a major concern, but security never becomes a major concern, the future system may be dominated by coal gasification and carbon capture with geological sequestration. **b** If, in contrast, gas prices first get lower, before then beginning to rise only after carbon control becomes more serious, one could imagine that a large number of additional central gas powered plants might come on line, followed by greater use of LNG imports and gas from coal gasification of coal. **c** If security first becomes a great concern without a rise in gas prices, one might see a much greater commitment to distributed co-generation and micro-grids. Then, as carbon prices start to rise, gas prices might also rise prompting more coal gasification and LNG. However, it would be hard to then convert distributed facilities to zero carbon emissions because H2 distribution is likely to be difficult. **d** Finally, if gas prices rise at the same time as security concerns grow, distributed generation might not grow as rapidly as in case **C**. Instead, one might see greater use of coal and later IGCC w/CCS and more attention to hardening the electric system against attack and providing back-up for carry-over in the event of outages

at the future moment of interest (e.g. trajectories A&B and trajectories C&D) while others might lead to very different end points (A&D versus C&D).

4.3 The use of dynamic models

The use of dynamic models in projections of energy futures presents a clear trade-off. On the one hand, models allow the integrative and systematic inclusion of dynamic features such as the demographic make up of the population, or primary energy supply. Moreover, models can enable use of the decomposition strategy described above to be applied to complex problems. On the other hand, use of models may obscure the crucial role of subjective judgments made by the modelers in choosing model structure and parameters. For example, they often do not treat the evolution of technologies and prices, in more than a very simple and deterministic way. Use of models should be parsimonious. The further we look into the future the less we can say with confidence. Models and scenarios should reflect that fact and become simpler as they are projected further into the future (Casman et al. 1999).

On time scales of the sort that were spanned by the SRES scenarios, formal models are probably best used as inputs to a broader process that weighs multiple sources of evidence. While this may include sensitivity studies of the values of uncertain parameters, as well as the structural form of a number of different models (Morgan et al. 2006), it should also include a variety of more qualitative considerations that are typically not captured, or at best captured inadequately, in formal models.

4.4 Probabilistic versus parametric analysis

One key design choice in the development of a projection is the selection of variables that will be treated probabilistically versus those that will be treated parametrically. We argued in Section 1 that to be effective projections must be shaped to meet the needs of specific decision makers. This typically means that variables over which the specific decision makers have no direct or indirect control should be treated probabilistically. However, variables over which those decision makers do have control should typically be treated parametrically so that the analysis provides insight about the implications of decisions or policy choices that could be made.

Thus, for national environmental policy makers a projection should treat the stringency of emission control parametrically so that they can see the consequences of choosing different values. Power company officials might also treat it parametrically if they are deciding what position to take in lobbying environmental policy makers. On the other hand, those same power officials would probably be best served by projections that assign probabilities to different levels of future emission control, based on their assessments of how regulations are likely to evolve, when they are faced with making major capital investment decisions about technologies with different emission profiles.

4.5 A strategy for developing probabilistic projections

By now it should be apparent to readers that developing an improved method for producing probabilistic projections (of energy use, CO_2 emissions, of similar quantities), for specific decision makers, that span periods of many decades, will require a significant research effort. We see no alternative but to make an informed judgment of what such an improved process might look like, based on the considerations we have discussed above; then try it out in some real context; subject the result to critical assessment from a wide range of perspectives; and then, refine the design and try again, informed by the results and the new insights that have been gained. In what follows, we sketch what we think is a reasonable design for a first effort.

Using all available models and other formal and informal tools and data at their disposal, a group of analysts (call them the assessment team) should first build as comprehensive a set of arguments for how the quantity of interest (call it Q) might end up at high, medium or low levels at the future moments of interest. This should be done by iterating back and forth between independent analysis by each member of the assessment team and collective discussion within the team.

Only after this process has iterated several times, should the team then move on to the task of imposing a parametric consideration of key policy variable(s) (P) and constructing a set of probability distributions in the value of Q at a future moment of interest, call it $p\,(Q\,(t = t_{\text{fut}})|\,P)$, a set of distributions at several times, or a continuous time series in $p(Q(t)|P)$.

Of course, different people may consider the detailed supporting work developed in the decomposition of the problem in order to produce different distributions. In most cases, it would be best to have several people do this independently and then compare, discuss, and perhaps iterate on the results. While at this point one might also want to involve some outside folks in making such judgments, the members of the assessment team who constructed the disaggregation should also engage in the process of probability assessment since Koehler (1994) has produced evidence that, having gone through the process of structuring the problem, they may be less overconfident and better calibrated in their judgments than outsiders who have not worked through the process of careful disaggregation.

It might also be advisable to adopt two different approaches to assessing the desired probability distribution(s). One could divide the range of plausible future values into five to seven intervals, and assess probabilities for each. In parallel, one could undertake a standard single probability assessment across the entire range (Morgan and Henrion 1990). First working as individuals, and then again after comparing notes with other respondents, each analyst could iterate back and forth between these two methods to develop their final distribution(s).

Because there is evidence that the number and content of categories across which probabilistic judgments are made may influence those judgments, it might also be wise in the first round of probabilistic judgments to let each respondent choose his or her own set of intervals across which to assess a probability distribution. On the one hand, we know that there is a strong tendency to overconfidence (i.e. to place too little probability in the tails of a distribution). On the other hand, the work of Fox and Clemen (2005) suggests that insufficient adjustment from "ignorance priors" may work toward a more uniform distribution of probability across the categories, in this case perhaps offsetting somewhat the impact of overconfidence.

While making judgments about the probability to attach to any range within the full distribution of interest, such as $p_{a<Q\leq b}\,(Q\,(t = t_{\text{fut}})\,|P)$ or $p_{a<Q\leq b}\,(Q\,(t)\,|P)$, clearly it will be important, to help team members and other participants consider the full range of outcomes collectively, so that the probability they assign to any subinterval, such as $a < Q < b$, is constrained by the need to have the total sum to 1.0. Similarly, it will be important to devise various cognitive aids to assist respondents in minimizing conjunction fallacies (Fox and Rottenstreich 2003; Tversky and Koehler 1994; Tversky and Kahneman 1983). The unpacking that was performed in the first stage of the process should be structured so that in this second stage analysts are making judgments across outcomes with "similar levels of specificity, rather than compar[ing] a single specific outcome against an unspecified set of alternatives" (Redelmeier et al. 1995).

Using such an approach it should be possible to obtain a set of carefully thought out individual probabilistic assessments. With additional effort and resources, it should also be possible to produce a group consensus probability estimate, using methods such as those that have been developed and demonstrated by Budnitz et al. (1995, 1998) in the context of assessing seismic hazard.

Computer based modeling tools may serve either as external inputs into this process or might play a more central role at a method of systematically aggregating judgments about various sub-processes where that aggregation is too complex to do manually. Modeling tools must be designed to incorporate probabilistic judgments and must be developed and communicated in a way that reveals rather than conceals the subjective judgments that are most important in driving the uncertainty in the outcome.

New information management and modeling tools may be needed. The development of the structured set of arguments in phase 1 is somewhat analogous to the development of features events and processes (FEP) catalogs developed in assessments of nuclear waste storage that are now being applied to assessments of CO_2 storage risks (OECD 1999, 2000). One might adapt similar tools to manage the catalog of assumptions and judgments used in the analysis proposed here.

Either way, those making probabilistic judgments should be fully versed in the literature on overconfidence and cognitive heuristics before they start the process. The literature on training to improve calibration in estimating probability distributions is not encouraging (Morgan and Henrion 1990). For example, Alpert and Raiffa (1982) reported little success in simply urging respondents to "spread those extreme fractals." Nonetheless, once a first iteration of probabilistic judgments has been completed, those making them should be asked to carefully consider the possibility that their distributions may be too tight, reflect on the poor performance of past projections, and, in that light, encouraged to think about possible revisions.

In at least some cases, the probability distributions that result from the process we have outlined may be much wider than decision makers would like. In other contexts, such as the application of downscaling, we have witnessed situations in which decision makers want "an answer" so that they can get on with the task at hand, and view large uncertainty as simply an annoyance to be avoided. There will probably always be decision makers who want to behave as though they know the future with greater precision than is possible. A careful analysis of the type we propose will not prevent such behavior, but it will certainly make it harder to justify.

4.6 Full probabilistic projections are not always needed

While this paper is focused on developing full probabilistic descriptions of the future evolution of energy use, carbon dioxide emissions, or similar parameters of interest, it is important to note that such projections are not needed to address some of the problems that decision makers face. For example, by working backwards from an outcome of interest, an analyst may be able to show that a decision is insensitive to an uncertain input parameter such as total energy use. Alternatively, the range of choices available to a decision maker may be sufficiently constrained that the value of an uncertain input makes no difference. Finally, a search for robust decisions (strategies that work well enough across the plausible range of outcomes), or the application of adaptive strategies (strategies that can be easily modified as the future unfolds), may not require a full probabilistic analysis. In all cases, analysts should let the problem they face drive the choice of methods they adopt.

5 Concluding thoughts

The motivation for a systematic exploration of energy futures is almost always to inform current decisions. The development of any projections of future energy use or emissions should be driven by the needs of the decision maker. Scenario developers may wish to provide general-purpose scenarios that would serve the needs of many decision makers. However, the heterogeneity of decision makers, the extraordinary complexity of possible futures, and the irreducible nature of many salient uncertainties, suggest that most such projections of the future cannot be both systematic and general.

Detailed story lines are often proposed on the grounds that they will help expand people's thinking. However, because they can be cognitively compelling, they may have quite the opposite effect, causing users to overlook a wide variety of alternate developments that could lead to similar outcomes for key variables such as energy use. Moreover, because they often describe a single point or line in large space of possible futures, it is not meaningful to attach probabilities to such scenarios. Developing a small number of very detailed story lines about the future is therefore not a promising way to support prospective analysis for climate and energy policy.

People's ability, operating holistically, to make successful long-range forecasts of key variables such as energy consumption, is generally poor. Moreover, many past estimates have shown strong systematic bias and considerable overconfidence. We, therefore, need a new approach to developing probabilistic inputs for use in long-term analysis for use by impacts and adaptation managers, energy resource and technology managers, government officials, and others responsible for developing regional and national policy or national positions in international frameworks.

We have argued that the development of such a new approach will require a serious research effort that iterates between the design of an improved process, applying and evaluating that process, and then refining the process and trying again.

We concluded by sketching what we believe would be a reasonable first design of an improved approach. Phase 1 would involve an iterative process of problem decomposition in which the objective would be to build as full a set of possible developments and factors that could influence the future value of the quantity or quantities of interest, including as complete a consideration as possible of path dependencies, legacy effects of capital stock, institutions and regulations, and similar issues. Then, in phase 2, a number of individual analysts, or perhaps a team of analysts, would move on to construct and refine probability distributions on the future values of the quantities of interest.

While there is certainly some room for commonality in projections that meet the needs of various decision makers, the choices of those quantities that should be projected probabilistically as opposed to the values that should be treated parametrically will depend on the needs of individual decision makers.

We believe that a method such as this should be capable of producing a much more useful set of inputs for use in climate and energy policy analysis and decision making, than the current approach of developing a few scenarios built on highly detailed story lines. It has the advantage that everyone who participates could help to develop and work from the same matrix of factors and developments that could influence future values. Different assessors might construct quite different probability assessments across those factors and developments. But the approach would allow assessors to more clearly explain and justify, and users to understand, the sources of those differences. Because of political sensitivities it might be difficult to implement such an approach as part of an official governmental or intergovernmental assessment process. However, done as a carefully performed and peer-reviewed independent effort, the results could be very valuable to many analysts as well as to official assessment efforts.

Acknowledgements Thanks to Garry Brewer, Michael DeKay, Baruch Fischhoff, Paul Slovic, and Joel Tarr for helpful advice and comments. This work was supported by the Climate Decision Making Center created through a cooperative agreement between the National Science Foundation (SES-0345798) and Carnegie Mellon University.

References

Alpert M, Raiffa H (1982) A progress report on the training of probability assessors. Harvard University. Reprinted. In: Kahneman D, Slovic P, Tversky A (eds) Judgment under uncertainty: heuristics and biases. Cambridge University Press, New York

Armstrong JS (1978) Long-range forecasting: from crystal ball to computer. Wiley Interscience, New York

Armstrong JS, Denniston WB, Gordon MM (1975) The use of decomposition principle in making judgments. Organ Behav Hum Perform 14:257–263

Arnell NW, Livermore MJL, Kovats S, Levy PE, Nicholls R, Parry ML, Gaffin SR (2004) Climate and socio-economic scenarios for global-scale climate change impacts assessments: characterizing the SRES storylines. Glob Environ Change 14:3–20

Bracken P (1977) Unintended consequence of strategic gaming. Simul Games 8:283–318 (September)

Bracken P (1990) Gaming in hierarchical defense organizations. In: Goldberg AC, van Opstal D, Barkley JH (eds) Avoiding the brink: theory and practice in crisis management. Brassey's, UK

Brewer GD, Shubik M (1979) The war game: a critique of military problem solving. Harvard University Press, Massachusetts

Budnitz RJ, Apostolakis G, Boore DM, Cluff LS, Coppersmith KJ, Cornell CA, Morris PA (1995) Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and the use of experts. Lawrence Livermore National Laboratory, UCRL-ID 122160

Budnitz RJ, Apostolakis G, Boore DM, Cluff LS, Coppersmith KJ, Cornell C, Morris PA (1998) Use of technical expert panels: applications to probabilistic seismic hazard analysis. Risk Anal 18:463–469

Casman EA, Morgan MG, Dowlatabadi H (1999) Mixed levels of uncertainty in complex policy models. Risk Anal 19(1):33–42

Craig P, Gadgil A, Koomey JG (2002) What can history teach US?: a retrospective examination of long-term energy forecasts for the United States. Annu Rev Energy Environ 27:83–118

de Finetti N (1974) Theory of probability: a critical introductory treatment. Wiley, New York

Edmonds J, Wilson T, Rosenzweig R (2000) A global energy technology strategy project addressing climate change: an initial report an international public–private collaboration. Joint Global Change Research Institute

EPRI (2007) The power to reduce CO_2 emissions: the full portfolio. Available at: http://eprireports.org/DiscussionPaper2007.pdf

Fischhoff B, Slovic P, Lichtenstein S (1978) Fault trees: sensitivity of estimated failure probabilities to problem representation. J Exp Psychol Hum Percept Perform 4:330–334

Fox CR, Birke R (2002) Forecasting trial outcomes: lawyers assign higher probability to possibilities that are described in greater detail. Law Hum Behav 26:159–173

Fox CR, Clemen RT (2005) Subjective probability assessment in decision analysis: partition dependence and bias toward the ignorance prior. Manage Sci 51:1417–1432

Fox CR, Rottenstreich Y (2003) Partition priming in judgment under uncertainty. Psychol Sci 14: 195–200

Goldberg AC, van Opstal D, Barkley JH (eds) (1990) Avoiding the brink: theory and practice in crisis management. Brassey's, UK

Greenberger M (1983) Caught unawares: the energy decade in retrospect. Ballinger, Cambridge

Gregory WL (2001) Scenarios and acceptance of forecasts. In: Armstrong JS (ed) Principles of forecasting: a handbook for researchers and practitioners. Kluwer, Massachusetts

Grübler A, Nakicenovic N (2001) Identifying dangers in an uncertain climate. Nature 412:15 (July 5)

Harte J (1985) Consider a spherical cow: a course in environmental problem solving. W. Kaufmann, CA

Henrion M, Fischhoff B (1986) Assessing uncertainty in physical constants. Am J Phys 54:791–798

Kahneman D, Slovic P, Tversky A (eds) (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, New York

Koehler DJ (1994) Hypothesis generation and confidence in judgment. J Exp Psychol 20:461–469

Lave LB, Epple D (1985) Scenario analysis. In: Kates RW, Ausubel JH, Berberian M (eds) Climate impact assessment. Wiley, New York

Lichtenstein S, Slovic P, Fischhoff B, Layman M, Combs B (1978) Judged frequency of lethal events. J Exp Psychol Hum Learn Mem 4:551–578

Lichtenstein S, Fischhoff B, Phillips LD (1982) Calibration of probabilities: the state of the art in 1980. In: Kahneman D, Slovic P, Tversky A (eds) Judgment under uncertainty: heuristics and biases. Cambridge University Press, New York

MacGregor DG (2001) Decomposition for judgmental forecasting and estimation. In: Armstrong JS (ed) Principles of forecasting: a handbook for researchers and practitioners. Kluwer, Massachusetts

MacGregor DG, Armstrong JS (1994) Judgmental decomposition: when does it work? Int J Forecast 10(4):495–506

Morgan MG, Henrion M, Small MS (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, New York

Morgan MG, Adams PJ, Keith DW (2006) Elicitation of expert judgments of aerosol forcing. Clim Change 75:195–214

Mullin TM (1986) Understanding and supporting the process of probabilistic estimation. Dissertation, Carnegie Mellon University

Nakicenovic N et al (2000) Special report on emissions scenarios. Cambridge University Press. Available online at: http://unep.no/climate/ipcc/emission/index.htm

OECD Nuclear Energy Agency (1999) Scenario development methods and practice, May

OECD Nuclear Energy Agency (2000) Features, events and processes (FEPs) for geological disposal of radioactive waste. An international database

Pacala S, Socolow R (2004) Stabilization wedges: solving the climate problem for the next 50 years with current technologies. Science 305:968–972

Parson E, Burkett V, Clarke L, Edmonds J, Fisher-Vanden K, Jacoby H, Keith D, Mearns L, Pitcher H, Reilly J, Richels R, Rosenzweig C, Webster M (2006) Scenarios of greenhouse gas emissions and atmospheric concentrations and review of integrated scenario development and application. US CCSP Synthesis and Assessment Product 2.1

Redelmeier DA, Koehler DJ, Liberman V, Tversky A (1995) Probability judgment in medicine: discounting unspecified possibilities. Med Decis Mak 15:227–230

Schneider SH (2001) What is 'dangerous' climate change? Nature 411:17–19

Schweizer VJ, Kriegler E (2007) Using cross-impact balance analysis to improve future emissions scenarios. Working paper. Carnegie Mellon University

Shell Group (2005) Shell Global Scenarios to 2025. Executive summary and excerpts available online at: http://www.shell.com/home/Framework?siteId=royal-en&FC3=/royal-en/html/iwgen/our_strategy/scenarios/dir_scenarios_28022005.html&FC2=/royal-en/html/iwgen/leftnavs/zzz_lhn5_4_0.html

Slovic P, Fischhoff B, Lichtenstein S (1976) Cognitive processes and societal risk taking. In: Carroll JS, Payne JW (eds) Cognition and social behavior. Halsted Press Division of Wiley, New York

Slovic P, Fischhoff B, Lichtenstein S (1980) Facts and fears: Understanding perceived risk. In: Schwing RC, Albers WA Jr (eds) Societal risk assessment: how safe is safe enough? Plenum, New York

Smil V (2003) Energy at the crossroads. MIT Press, Massachusetts

Tarr J (1996) The search for the ultimate sink: urban pollution in historical perspective. University of Akron Press, Ohio

Tversky A, Kahneman D (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol Rev 90:293–315

Tversky A, Koehler DJ (1994) Support theory: a nonextensional representation of subjective probability. Psychol Rev 101:547–567
US National Assessment (2001) Climate change impacts on the United States—foundation report. Cambridge University Press, New York
Wigley TML, Raper SCB (2001) Interpretation of high projections of global-mean warming. Science 293:451–454
Yergin D (1991) The prize: the epic quest for oil, money and power. Free Press, New York