

Reducing doubt about uncertainty: Guidance for IPCC's third assessment

Richard H. Moss

Received: 15 May 2011 / Accepted: 19 July 2011 / Published online: 9 September 2011
© U.S. Government 2011

Abstract It is usually necessary to apply incomplete and uncertain information to inform policy and decision making, creating the need to characterize the state of knowledge and identify when more certain information may be available. After all, some information is better than none and conversely, even perfect information is of no use if it is available only after a decision has been made. In scientific assessments for global change, the challenges are particularly acute because of scientific complexity, long time horizons, and large political and economic stakes, among other factors. Moss and Schneider prepared uncertainty guidelines for the Third Assessment Report (TAR) of the Intergovernmental Panel on Climate Change (IPCC) that recommended a process to make expert judgments of levels of confidence and uncertainty more systematic and transparent. The guidance provided calibrated uncertainty terms to improve communication of findings to users and urged preparation of a traceable account of the authors' assessment of the evidence for each major finding. This article reviews the recommendations and their effectiveness and highlights ensuing critiques and the evolution of uncertainty guidance for subsequent assessment reports. It discusses emerging challenges in providing science for decision making in the era of increasing model resolution and complexity and burgeoning interest in information to inform adaptation and mitigation at regional and finer scales.

“Uncertainty, or more generally, debate about the level of certainty required to reach a ‘firm’ conclusion, is a perennial issue in science. The difficulties of explaining uncertainty become increasingly salient as society seeks policy prescriptions to deal with global environmental change. How can science be most useful to society when evidence is incomplete or ambiguous, the subjective judgments of experts about the likelihood of outcomes vary, and policymakers seek guidance and justification for courses of action that could cause significant societal changes? How can scientists improve their characterization of uncertainties so that areas of slight disagreement do

The author is Senior Staff Scientist at Pacific Northwest National Laboratory's Joint Global Change Research Institute at the University of Maryland. He dedicates this article to the memory of Stephen H. Schneider, the captain of the “uncertainty cops”. He also acknowledges with gratitude PNNL colleagues Jennie Rice, Michael Scott, and Stephen Unwin, whose knowledge and ideas have given shape to a number of recommendations in this article. He also thanks Elizabeth Malone, Granger Morgan, Nathan Engle, and Caroline Boules for comments on an earlier draft.

R. H. Moss (✉)

Joint Global Change Research Institute, University of Maryland, Pacific Northwest National Laboratory,
Maryland, USA
e-mail: rhm@pnnl.gov

not become equated with purely speculative concerns, and how can individual subjective judgments be aggregated into group positions? And then, how can policymakers and the public come to understand this input and apply it in deciding upon appropriate actions? In short, how can the scientific content of public policy debates be fairly and openly assessed?" (Moss and Schneider 2000)

These questions were raised in the opening paragraph of the synthesis essay for an Aspen Global Change Institute workshop held during the summer of 1996 and seem as relevant today as when they were posed. The issues motivated development of the guidance document on uncertainty for authors of the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report (TAR), an early effort to apply the state of the art in decision analysis and risk communication to climate assessment. The guidance built on the existing literature (e.g., Funtowicz and Ravetz 1990; Morgan and Henrion 1990) and experience gained in a small number of previous expert elicitations (e.g., NDU 1978; Nordhaus 1994; Morgan and Keith 1995; Titus and Narayanan 1995) and included specific recommendations for a process to make subjective judgments of lead author teams more systematic and comparable by, among other things, use of common calibrated language for describing likelihoods and levels of confidence.

But the calibrated language was just a small part of the recommendations. The main focus was a process designed to increase attention to uncertainties and confidence and help lead author teams avoid some of the common cognitive traps that the literature highlighted. Improving dialog and understanding across disciplines from the natural and social sciences and engaging specialists in the field of decision analysis were additional key objectives because IPCC assessments needed to integrate findings on emissions of greenhouse gases, changes in radiative forcing, climate system response, implications for impacts, and options for reducing emissions. The guidance also addressed challenges in communicating results to users identified in evaluation of prior reports that indicated that lay persons (e.g., in the media and policy communities) would not accurately interpret assessment findings without a clear approach and language that enabled them to distinguish what is speculative from what is well established.

This article will revisit the context of the first version of uncertainty guidance for the IPCC, review the main elements of the recommendations, mention some of the critiques and subsequent improvements in versions of the guidance prepared for the Fourth and Fifth Assessment Reports (AR4 and AR5, respectively), and suggest some next steps for further improvements. Other articles in this special issue provide a more detailed analysis of the AR4 and AR5 guidance.

1 Assessing science for public policy and decisionmaking

Climate change belongs to a class of problems that are noted for their scientific complexity and ability to generate political controversy. The decision stakes are high—trillions of dollars in investments in energy infrastructure, geopolitical issues and trade flows, the environmental heritage and resources we leave for future generations, and the viability of Earth's planetary life support systems. The issues are complex, involving interactions and feedbacks among multiple human and natural systems, e.g., population and economic development, ecosystems and species diversity, hydrology, food and fiber production, nutrient cycles, and the climate system itself. And the environmental harm is not obvious to casual observers, at least not yet, because the problem is largely caused by odorless, colorless gases whose immediate effects can be difficult to observe in the context of weather and seasonal to interannual climate variability.

Using uncertain evidence to inform high-stakes choices is not a novel challenge. The intelligence community grappled in the early 1960s with efforts to quantify what were

essentially qualitative judgments and apply calibrated uncertainty terminology to improve intelligence reports (CIA 1964), paving the way to development of techniques which are now used in describing probabilities and levels of confidence in national intelligence estimates (NIC 2007). The medical community has sought to communicate potential harms and benefits of medical interventions in shared medical decisionmaking among patients, physicians, and policymakers (Politi et al. 2007). Other examples include management of radiological hazards, approval of waste management sites and technologies, drug testing and approval, power plant siting, land use planning, water resources development, regulation of potentially carcinogenic chemicals, and more.

In the case of climate change, the challenges of applying science to inform decision-making are particularly acute, in part because models play such a central role in attributing observed changes to human activities and projecting the future evolution of climate over decades to centuries. According to Pidgeon and Fischhoff, “much climate science relies on simulation modelling that is an unfamiliar form of inference not just for lay people but even for scientists whose disciplines use observational methods. Unless the logic of that modelling is conveyed, people may discount its conclusions” (Pidgeon and Fischhoff 2011). Conflicting model results, different subjective interpretations of those results by different groups of experts, and the absence of methods for aggregating different expert characterizations of uncertainty have contributed to lack of progress in policy formulation, at a minimum by providing an opportunity for special interests to confuse and divert public discourse.

In the policy debate about climate change, uncertainty is often approached as a generalized issue—as an overall reason for inaction until science becomes increasingly certain and thus provides the basis for action. Unfortunately, uncertainties will persist long after specific adaptation and mitigation decisions need to be made. Confusion and drawn out debates could be avoided if uncertainty were considered in the context of a particular decision (or science question) because it is this context that dictates the level of certainty required—understanding the sign or direction of a potential change might be adequate for some contexts but fall completely short for others.

The need to apply incomplete and uncertain information at a point in time to a specific decision creates the need for approaches to characterize the state of knowledge and identify when more certain information may be available. The IPCC reports, and scientific assessments more generally, are intended to meet this need (Agrawala 1998a; Agrawala 1998b), making it essential that IPCC accurately portray and interpret the range of results in the literature and characterize the level of confidence in findings in order to inform policy. Improving IPCC’s assessment capacity was the motivation for Steve and me to develop the uncertainty guidance for the TAR. Each of us had been involved in previous IPCC assessments—Steve as a lead author and I as the head of the technical support unit for Working Group II, which at the time focused on impacts, adaptation, and technological aspects of mitigation. Each of us had independently attempted to develop some consistency and transparency in how different teams of lead authors conveyed to audiences the extent of certainty in their conclusions. We had seen the difficulties within the author teams in characterizing the scientific and technical aspects of uncertainty, for example when authors reached different conclusions based on their subjective interpretations of results, what Patt would later describe as the distinction between model-based and conflict-based uncertainty (Patt 2007). And we had noted the miscommunication that occurred between the scientists in the assessment and users because decision context was not considered and the language employed was inconsistent and at times inaccessible. It

was clear that something needed to be done to address both the uncertainty *characterization* problem as well as the deficiencies in *communication* of confidence and uncertainty.¹

In spite of the obvious central role of uncertainty characterization in assessment and the need to make such judgments as systematic and transparent as possible, the initial suggestion to prepare the guidance and disseminate it to lead authors of the TAR was not welcome in all quarters of the IPCC. A number of authors, especially among climate scientists, were uncomfortable. As noted in a later review, "...Many authors of the TAR were unused to quantifying their subjective judgments in terms of numeric probabilistic ranges. Physical scientists, in particular, prefer to consider uncertainty in objective terms, typically in relation to measurements and as something that can be estimated from the repeatability of results and by determining bounds for systematic errors" (Manning 2003). In addition, some IPCC leaders and government representatives who participated in the IPCC process worried that focusing on uncertainty would reduce the legitimacy of the reports for decisionmakers by making the knowledge seem limited and conditional. Another reviewer of IPCC's efforts to characterize uncertainty subsequently noted that "In the IPCC process, political and epistemic motives can be found to be intertwined, sometimes leading to the suppression of uncertainty communication ... the public expression of 'intra-peer community differences' is subdued due to the presence of greenhouse skeptics in society, who are typically very vocal critics of the IPCC" (Petersen 2006). Fortunately, the view that decision-making and science were better served by making treatment of uncertainties more explicit trumped these concerns, leading to development and application of guidance in all subsequent IPCC reports.

2 Uncertainty guidance for the third assessment report

The process of preparing the guidance actually began outside the IPCC itself, with a nine-day research workshop held during August 1996 at the Aspen Global Change Institute. The workshop brought together statisticians, decision analysts, media and policy analysts, and a number of lead authors from the IPCC Second Assessment Report (SAR) who spanned the physical, biological, and social sciences. The goal of the workshop was to review prior experience in the IPCC and identify methodological options for improving the evaluation and compositing of judgments of researchers.

Workshop participants tackled tough issues, a mix of theory, methods, and empirical problems that arose in a sample of key issues from the three IPCC working groups. These issues included climate sensitivity, global mean surface temperature increase, attribution of climate change to anthropogenic causes, impacts of climate change on vegetation-ecosystems and agriculture-food security, and estimation of aggregate economic impacts of climate change and emissions abatement. The workshop reviewed a variety of methods,

¹ Approaches to evaluating and communicating uncertainty in expert-judgment based scientific assessments are a small subset of quantitative and qualitative approaches to uncertainty quantification, propagation and characterization applied to evaluating uncertainties in research and analysis. Uncertainty characterization and communication is essential to ensure that a scientific conclusion represents a valid and defensible insight. This article does not address this broader set of approaches for evaluating uncertainty in the science of complex systems.

including approaches to expert elicitations and using an “assessment court” in which independent judges provide a neutral evaluation of the assessment of a set of lead authors.

The workshop also reviewed how different scientific/technical debates or changes in conclusions from the previous IPCC assessment played out in media coverage and policy debates. One example examined how the SAR’s 1°C lower estimate of global mean surface temperature increase (compared to conclusions in the first IPCC assessment) was misinterpreted by some as an indication that the climate sensitivity was lower than previously thought, when in fact the reduction was due to the addition of sulfate aerosols (which have a cooling effect) in the emissions scenarios and the GCMs. The workshop results provided the foundation for the TAR uncertainty guidance.

2.1 A systematic process

The TAR uncertainty guidance included specific recommendations for a process that was intended to make subjective judgments of lead author teams more systematic and comparable by raising awareness of issues that could affect the authors’ deliberations and presentation of results. The guidance pointed out that while science strives for “objective” empirical information, assessment involves being responsive to a decisionmaker’s need for information at a particular point in time, given the information available, and that assessing the quality of existing information involves a large degree of subjectivity. The guidance was intended “to make such subjectivity both more consistently expressed (linked to quantitative distributions when possible) across the TAR, and more explicitly stated so that well established and highly subjective judgments are less likely to get confounded in policy debates” (Moss and Schneider 2000).

The overview stated that “care should be taken to avoid vague or very broad statements with ‘medium confidence’ that are difficult to support or refute.... The point is to phrase all conclusions so as to avoid nearly indifferent statements based on speculative knowledge. In addition, all authors—whether in Working Group I, II or III—should be as specific as possible throughout the report about the kinds of uncertainties affecting their conclusions and the nature of any probabilities given.” And finally, the overview admonished authors to clearly describe their approach and assumptions—“Transparency is the key in all cases” (Moss and Schneider 2000).

The guidance document outlined a process to be followed by the lead authors in “a relatively small number of the major conclusions and/or estimates (of parameters, processes, or outcomes) It is not intended that authors must follow all of these steps every time they use a term such as ‘likely’ or ‘unlikely’ or ‘medium confidence’ in the main text of their chapters or every time a specific result is given.” Moreover, the guidance was never intended to be a “one size fits all” solution for all the issues addressed across the three IPCC working groups. “Given the diverse subject areas, methods, and stages of development of the many areas of research to be assessed in the TAR, the paper cannot provide extremely detailed procedures that will be universally applicable. Therefore, this document provides general guidance; writing teams will need to formulate their own detailed approaches for implementing the guidance while preparing their chapters and summaries” (Moss and Schneider 2000).

The process involved a set of six steps (and an optional seventh that we hoped might be applied to especially controversial topics such as attribution) that are summarized verbatim in Box 1. Readers who wish to investigate the details of each of these steps should consult the guidance itself, but several points are worth emphasizing. A key aspect of the guidance was the order in which these steps were recommended, a process intended to guard against a number of heuristic shortcuts often employed by experts and lay audiences alike in making judgments under uncertainty (Morgan and Henrion 1990). Beginning this process

Box 1: Summary of steps recommended for assessing uncertainty in the TAR

1. For each of the major findings you expect to be developed in your chapter, identify the most important factors and uncertainties that are likely to affect the conclusions. Also specify which important factors/variables are being treated exogenously or fixed, as it will almost always be the case that some important components will be treated in this way when addressing the complex phenomena examined in the TAR.
2. Document ranges and distributions in the literature, including sources of information on the key causes of uncertainty. Note that it is important to consider the types of evidence available to support a finding (e.g., distinguish findings that are well established through observations and tested theory from those that are not so established).
3. Given the nature of the uncertainties and state of science, make an initial determination of the appropriate level of precision—is the state of science such that only qualitative estimates are possible, or is quantification possible, and if so, to how many significant digits? As the assessment proceeds, recalibrate the level of precision in response to your assessment of new information.
4. Quantitatively or qualitatively characterise the distribution of values that a parameter, variable, or outcome may take. First identify the end points of the range that the writing team establishes, and/or any high consequence, low probability outcomes or “outliers.” Particular care needs to be taken to specify what portion of the range is included in the estimate (e.g., this is a 90% confidence interval) and what the range is based on. Then provide an assessment of the general shape (e.g., uniform, bell, bimodal, skewed, symmetric) of the distribution. Finally, provide your assessment of the central tendency of the distribution (if appropriate).
5. Using the terms described below, rate and describe the state of scientific information on which the conclusions and/or estimates (i.e. from step 4) are based.
6. Prepare a “traceable account” of how the estimates were constructed that describes the writing team’s reasons for adopting a particular probability distribution, including important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, and critical uncertainties.
7. OPTIONAL: Use formal probabilistic frameworks for assessing expert judgment (i.e. decision analytic techniques), as appropriate for each writing team.”

with identification of key uncertainties and documentation of the ranges of estimates found in the literature was intended to sensitize the authors to the factors that would make their conclusions uncertain. Consideration of the appropriate level of precision given the available evidence (and we should have added the context of the potential application of the information) would help authors avoid what Steve was fond of referring to as “crack-pot rigor”. And in the process of characterizing distributions, starting by identifying end points of probability distributions and possible outliers, and then developing estimates of the central tendency only if warranted, would help authors to avoid anchoring their estimates on the central tendency and thus misleading users of the report by reporting too narrow a range of possible outcomes.

The process of characterizing distributions received a great deal of attention in the guidance because of the implications of providing truncated ranges on the ability of the report to convey uncertainties accurately. Some authors were uncomfortable with including the full range because the likelihood of a “surprise” or events at the tails of the distribution could be too remote to gauge. The guidance encouraged inclusion of such surprises, if not by describing the full range then by providing a truncated range in addition to outliers”. The guidance encouraged authors to provide an assessment of the shape of the distribution, allowing for uniform distributions, noting that under these circumstances, writing teams may not consider it appropriate to make a “best guess”. It pointed out the importance of not combining different distributions that result from different interpretations of the evidence or “schools of thought” into one summary distribution.

In step 5, the guidance suggested two sets of terms for rating and describing the “quality of scientific information/level of confidence” in the conclusions or estimates (Figs. 1 and 2). The point was to encourage comparability of the way that level of confidence in results was described across the three IPCC working groups. The importance of this step was evident in the observation that without some sort of calibration, readers and even the authors of the report interpreted terms such as “probable,” “likely,” “possible,” “unlikely,” and “doubtful” very differently—for example respondents associated a surprisingly broad range of probability (0.6–1.0) with the qualitative term “almost certain” (Morgan 1998). We argued that consistency was needed for a number of reasons. First, common vocabulary would reduce linguistic ambiguity. Second, common, calibrated terminology would help the assessment communicate uncertainty to policymakers and the public and thus increase the accuracy of interpretation and utility of the assessment by users. Finally, because the IPCC was asked by the UNFCCC to address a broad range of cross-cutting questions related to its deliberations, it was necessary to synthesize results from across the three working groups. While specialists in one field might be aware of language used to describe uncertainty and level of confidence in their field, they were much less familiar with practice in other fields (a condition that persists today). A common confidence vocabulary was that it would facilitate compositing of uncertainty in complex outcomes that span disciplines and working groups.

The decision to use words to describe level of confidence was based on evidence that this is an intuitive approach for most people. By linking these words to a simple quantitative scale, we hoped that ambiguity in interpretation of the terms would be reduced (Fig. 1). As a result of the email reviews and teleconferences of lead authors critiquing early drafts of the guidance, we knew that some authors were uncomfortable with having only one option for describing confidence. Thus we proposed a supplementary set of qualitative uncertainty terms to describe why a writing team might express a particular level confidence in a finding (Fig. 2). The qualitative terms were more explicit in considering

(1.00)
“Very High Confidence”
(0.95)
(0.95)
“High Confidence”
(0.67)
(0.67)
“Medium Confidence”
(0.33)
(0.33)
“Low Confidence”
(0.05)
(0.05)
“Very Low Confidence”
(0.00)

Fig. 1 Scale for assessing state of knowledge

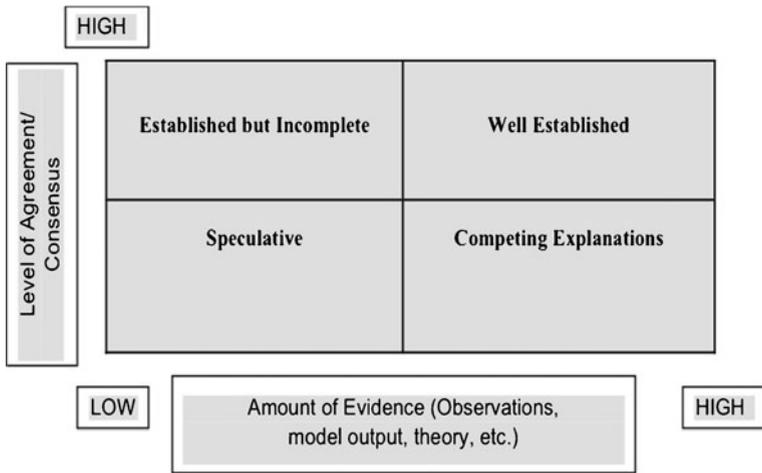
two of the factors that could bear on level of confidence: the amount of information that was available and the degree of consensus among relevant experts in the interpretation of that information. We also proposed potential graphical approaches for communicating degree of confidence, including a radar plot (Fig. 3).

Finally, as indicated above, we felt the most important step was to provide a transparent and traceable account of how judgments were reached. The guidance highlights the importance of this and suggests the account should include “important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, and critical uncertainties,” giving among other examples a hypothetical case in which outliers are added to ranges based on model output because the model doesn’t incorporate specific processes that would, in the judgment of the authors, increase the range.

3 Results and critiques

The guidance sparked a great deal of commentary and debate, and in this respect it succeeded in at least one of its aims, raising the profile of the issue. According to Manning and Petit, both of whom contributed to the TAR and were involved in preparing the revision of the guidance for the AR4, “... the approaches adopted by different groups of authors were similar in many cases and some advances were achieved over the Second Assessment.... For example, the use of doubly caveated statements of the form ‘we have medium confidence that phenomenon X might occur’ was largely avoided....” (Manning and Petit 2003). Patt and Dessai concluded “In general, the approach of the IPCC TAR should be considered a step in the right direction” (Patt and Dessai 2005).

Most critiques noted the guidance didn’t achieve all it set out to. Dessai and Hulme observed that the guidance was “halfheartedly followed by the various chapters of each Working Group, with some conforming to the framework more closely than others” (Dessai and Hulme 2004). A simple word count of the contributions of the three Working Groups to



Well-established: models incorporate known processes; observations consistent with models; or multiple lines of evidence support the finding)

Established but Incomplete: models incorporate most known processes, although some parameterizations may not be well tested; observations are somewhat consistent but incomplete; current empirical estimates are well founded, but the possibility of changes in governing processes over time is considerable; or only one or a few lines of evidence support the finding

Competing Explanations: different model representations account for different aspects of observations or evidence, or incorporate different aspects of key processes, leading to competing explanations

Speculative: conceptually plausible ideas that haven't received much attention in the literature or that are laced with difficult to reduce uncertainties

Fig. 2 Supplemental qualitative uncertainty terms

the TAR confirms the differences among the three groups, and led Swart et al. to conclude that each working group not only used different terms, but even fundamentally different framings of uncertainty (for details see <http://www.centre-cired.fr/forum/article428.html>) (Swart et al. 2009).

In the spirit of providing highlights but not an exhaustive review of the articles spawned by the guidance, I will group comments on specific aspects of the guidance into the following categories: (1) failure to communicate with intended audiences; (2) failure to harmonize the “confidence” language or develop a clear approach to “likelihoods” or probabilities; (3) inappropriately forcing one uncertainty characterization process onto three very different epistemologies; and (4) process problems that contributed to the application of the approach in the different working groups.

3.1 Communication failures and disputes over terminology

While appealing in its simplicity, a number of authors have argued that the intuitive approach of using calibrated uncertainty terms taken in the guidance failed to improve communication with the audiences of the report. The issue of audience is both complex and important; the approach one chooses to communicate a finding and the degree of confidence in it must be tailored to the recipients. IPCC reports have a wide set of

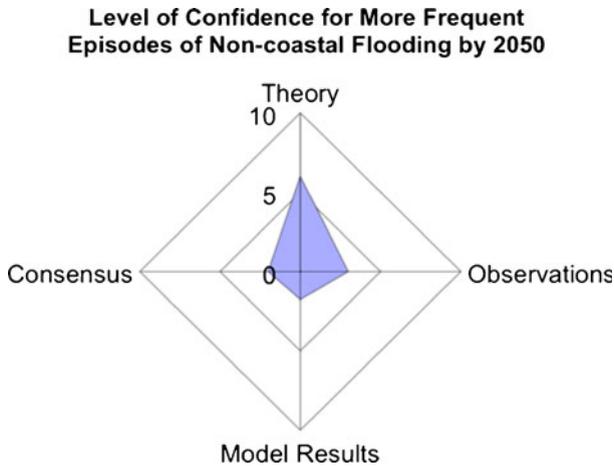


Fig. 3 Graphical representation of multiple dimensions of evidence that could contribute to confidence in a finding (a discrete outcome) using a simple 10-point scale for each of four dimensions

audiences, including delegates to the UNFCCC, policymakers in national governments, officials in regional organizations, non-governmental organizations and businesses, and for the technical chapters, subject matter experts working in areas of resource management such as water, forestry, or agriculture. Thus the ‘appropriate’ approach to communication could vary from qualitative terms to probability density functions.

The approach used in the guidance, attempting to calibrate individuals’ intuitive understanding of specific terms, held the promise of providing a minimum common denominator for these audiences, with options to add more sophisticated approaches for more technical groups. Subsequent research has demonstrated that use of this intuitive approach is subject to a variety of biases, including ambiguity, probability weighting towards the center of a distribution, and context dependency. Experiments conducted with Boston University science students and participants in UNFCCC negotiations indicated that most people’s intuitive approach to likelihood terms was not influenced by the definitions provided by the IPCC. These studies indicate that lay interpretation incorporated both the probability and the magnitude of the risks, whereas IPCC’s use of the terms focused simply on the probability. Even with clarifications of the distinction, they concluded that policymakers’ interpretation “is likely to be biased downward, leading to insufficient efforts to mitigate and adapt to climate change” (Patt and Schrag 2003; Patt and Dessai 2005). This finding has implications not only for communication, but for the overall approach to “risk” in a situation in which the expertise to assess likelihoods of changes in the climate system resided in one WG, and expertise to assess the magnitude of the damage these changes might produce resides in another (Manning 2003). A more recent experiment indicates that the intuitive approach is not as effective as hoped in calibrating interpretation of results by users (Budescu et al. 2009) and has led to recommendations to use simple numerical approaches instead of calibrated qualitative terms (e.g., Pidgeon and Fischhoff 2011).

Another criticism of the uncertainty terminology was a problem with providing seemingly precise ranges of probability to be associated with specific terms. The point was that as presented, the calibration scale seemed to indicate that experts intended for there to be a sharp distinction in level of confidence between statements judged to have a 33%

likelihood of being correct (the top of the range for low confidence) from those with a 34% likelihood (the bottom of the range for medium confidence). Such implied precision was unjustified and confusing, leading to the suggestion of using fuzzy boundaries, as was subsequently done in the US National Assessment (USGCRP 2000).

3.2 Failure to harmonize confidence terms and underlying approaches to likelihood and probability

Several articles have focused on the fact that the guidance did not actually result in harmonization of language across the report. The guidance took a very simple approach—modifying the word “confidence” with adjectives from “very high” to “very low” in a five-point scale. The intended interpretation of these terms was that the authors had confidence in the indicated range (e.g., 95 to 67% chance) that the statement was correct. In preparing its contribution to the TAR, WG I decided to use its own seven level scale of *likelihoods* of outcomes, from virtually certain (“greater than 99% chance that a result is true”) to exceptionally unlikely (less than 1% chance). Two motivations have been offered for use of this alternative: to extend the scale to seven levels to provide terms for statements that seemed almost certain or extremely unlikely (Swart et al. 2009) and to provide language that was more consistent with the frequentist² preferences of some WG I authors (Petersen 2006). Petersen points out that the introduction of the likelihood scale resulted in some lead authors using it in a frequentist mode and other authors (and the SPM) stating that the scale represented subjective judgments. Examining one of the most consequential findings of the report, on the attribution of observed changes to human causes, he notes that both frequentist and subjective approaches were combined to produce the finding that “*most* of the observed warming over the last 50 years is *likely* to have been due to the increase in greenhouse gas concentrations” (emphasis added) (IPCC 2001a). “He concludes “the ‘likelihood’ terminology cannot adequately represent model unreliability. At least, it is difficult if not impossible to distill the lead authors’ judgement of climate-model unreliability, as it influences the attribution conclusion, from the word ‘likely’” (Petersen 2006).

The introduction of the second likelihood scale by WG I meant that uncertainty or confidence rankings would not be used clearly and consistently across the TAR, with especially serious consequences for the Synthesis Report (IPCC 2001d). Box SPM-1 listed the scales from both the guidance and WG I and indicated that the likelihood scale applied to findings from WG I (IPCC 2001a), the confidence scale from the guidance applied to WG II (IPCC 2001b), and no confidence levels were assigned to WG III (IPCC 2001c). From a reading of Box SPM-1, it is difficult to understand how an average non-technical user could make sense of this profusion of descriptors, or how one could arrive at confidence statements regarding conclusions that drew on findings from all three groups. Additional confusion was sewn when, in response to the final question addressed in the Synthesis Report, the additional terms “robust” and “key uncertainties” were introduced and defined as follows: “In this report, a *robust finding* for climate change is defined as one that holds under a variety of approaches, methods, models, and assumptions and one that is expected to be relatively unaffected by uncertainties. *Key uncertainties* in this context are

² In the frequentist statistical approach, probability is defined as the relative frequency that something would occur in a limitless number of independent, identical trials. In practice, random sampling or other thought experiments are used since there are relatively few phenomena that can be expected to occur in an infinite, independent, and identical fashion. A Bayesian approach involves a rigorous mathematical methodology for evaluating the probability of an event conditional on the available information, interpreted as degree of belief, which is updated as new evidence becomes available.

those that, if reduced, may lead to new and robust findings in relation to the questions of this report” (IPCC 2001d). In the subsequent tables, a number of findings were listed as “robust” with the use of qualifiers such as “most”, “some”, and “substantial” that made the statements indifferent. These are prominent and frustrating examples of some of the very problems we sought to avoid with the uncertainty guidance.

Underlying WG I’s decision to use their own likelihood scale was the critique that the guidance did not distinguish between likelihood and level of confidence and lacked an explicit approach for treating frequentist statistical claims (Petersen 2006). This issue was explored at length in a preparatory workshop for the AR4 convened in Maynooth, Ireland to review the TAR experience and encourage debate about different ways of characterizing and communicating uncertainty. In a conference paper for this workshop, Allen and colleagues noted “we need to communicate the fact that we may have very different levels of confidence in various probabilistic statements” (Allen et al. 2004). These authors proposed to first use an objective likelihood assessment for the first stage followed by a subjective confidence assessment, defined as degree of agreement or consensus. Contradicting this point, others argue that likelihood and confidence cannot be fully separated. Likelihoods, especially extremely likely or unlikely ones, contain implicit confidence levels. “When an event is said to be extremely likely (or extremely unlikely) it is implicit that we have high confidence. It wouldn’t make any sense to declare that an event was extremely likely and then turn around and say that we had low confidence in that statement” (Risbey and Kandlikar 2007). This issue was addressed in different ways in the subsequent guidance documents for the AR4 and AR5.

There was no coordinated approach to assessing probabilities, and thus the TAR failed to estimate probabilities of different levels of climate change for different periods of the 21st century. A number of papers explored this issue and included criticisms and rebuttals from lead authors (e.g., Allen et al. 2001; Reilly et al. 2001; Schneider 2001). Steve’s commentary in *Nature* arguing that the Special Report on Emissions Scenarios should have estimated likelihoods for its emissions scenarios to provide the basis for such estimates (Schneider 2001) sparked an extremely lively debate (Grubler and Nakicenovic 2001; Lempert and Schlesinger 2001; Pittock et al. 2001). At its most basic level, the debate involved differences of opinion about whether assigning a probability distribution for future emissions required making predictions for unpredictable human decisions. The Working Group I contribution to the AR4 noted that the TAR’s “failure in dealing with uncertainty [in] ... the projection of 21st-century warming ... makes the interpretation and useful application of this result difficult” (IPCC 2007). The debate continues to evolve and a large literature (too extensive to review systematically here) has developed to formulate probability distributions of future climate change using a variety of techniques including fuzzy sets, possibility theory, Bayesian model averaging, and others (e.g., Ha-Duong 2003; Krieglger and Held 2005; Hall et al. 2007; Min et al. 2007).

3.3 Forcing one size to fit all

Some have argued that the attempt to develop one consistent approach to uncertainty characterization across the IPCC Working Groups ignored fundamental differences in the nature of uncertainty in different areas of research. For Working Group I, uncertainties in the physical sciences related to empirical quantities observed in the natural world (for which there can arguably be said to be a “true” value) were among the issues considered, and physical scientists preferred to consider these issues in objective terms using frequentist approaches. In Working Group II, with a focus on observed and projected impacts, data

were usually less precise, and uncertainties resulted from a range of sources related to observations, models, and differences in perspective regarding the value of impacts, and authors were more comfortable with a subjective approach, as called for in the guidance. For Working Group III authors, who focused on such topics as emissions, mitigation potentials, and costs and benefits, many of the main uncertainties concerned human decisions, and as a result “the authors opted for addressing uncertainty differently, mainly through presentation of ranges of outcomes and explanation of the underlying assumptions.”

There are many different typologies of uncertainty. In some fields, taxonomies of uncertainty types and sources have become stable so that techniques for uncertainty characterization are standardized for different uncertainty sources. For climate science, no such standardized typology has been established with the result that many different typologies focus on the needs of different subfields. Across these there are some common elements, with many authors differentiating aleatory uncertainty of climate’s internal variability, epistemic uncertainties associated with unknowns in how the climate system will respond to anthropogenic emissions (represented, *inter alia*, in inter-model differences), and uncertainties resulting from human agency and decisions (e.g., the factors producing uncertainty in potential future emissions). Distinguishing different sources of uncertainty is important, as these diverse causes have varying effects on projections for different variables and time periods (e.g., Hawkins and Sutton 2010) and it is important that decisionmakers have some understanding of which are important for different circumstances and issues. But it is debatable whether these differences completely blur the distinctions in classical concepts of measurement theory and error classification and thus require completely different approaches for propagation, quantification, or qualitative characterization as has been argued (e.g., Swart et al. 2009). For example, there are a variety of standard approaches that can be used to represent uncertainty in human decisionmaking, even if probability distributions for those variables are not developed.

The TAR guidance did not set out to develop a single universal approach to uncertainty in research related to climate change. As discussed earlier, it did seek to establish a least common denominator for key findings while encouraging development of methods better suited to specific circumstances. Nevertheless, Swart and colleagues concluded that because the TAR guidance blurred important distinctions in the sources of uncertainty, in the future, a diversified approach should be followed in which the Working Groups are freer to use methods most appropriate to the uncertainties in their domains (Swart et al. 2009). However, Steve and I felt—and I still feel—that while diverse approaches are needed for different issues, it would be a mistake to abandon this quest for some level of consistency, especially to support synthesis. In its recent review of IPCC procedures and methods, the InterAcademy Council reaffirmed support for a unified approach, stating that all three working groups should use a qualitative level-of-understanding scale in their policymaker and technical summaries and only use the likelihood scale or quantify probability if the outcomes are well-defined and there is sufficient evidence (IAC 2010).

3.4 An under-staffed police force and other implementation issues

Finally, a number of practical issues affected efforts to implement the uncertainty guidance, including inadequate resources, lack of time and other priorities, and finalization of the guidance too late in the process. Reilly and colleagues asserted that the process started too late in the TAR, noting that “uncertainty analysis should not be pasted on to the end of an assessment, but needs to be implemented from the beginning...” (Reilly et al. 2001). While some of the initial working group lead author meetings did introduce the draft guidance, it

is true that the guidelines were not finalized until the process was well underway, making it more difficult to establish consistency in approach than it might otherwise have been.

Steve and I were dubbed “the uncertainty cops” for our efforts to encourage application of the guidance (Giles 2002), but in fact there were simply too few individuals who were trained in decision analysis or were familiar with the details of the guidance to actually team with lead authors in implementing the guidance in their chapters. It is a big job: in total, the three Working Group reports contained 43 chapters, plus additional technical and policymaker summaries, and the TAR also included an extensive “Synthesis Report” that addressed nine cross-cutting questions. A small army of experts would be needed to cover even the three to four most important conclusions across this extensive set of materials. The lack of individuals with assigned responsibility for “policing” the uncertainty guidance was noted in a concept note preparing for the AR4, along with a proposal to identify reviewers drawn from outside the normal pool of climate experts to focus on the treatment of uncertainty in the WG reports (Manning and Petit 2003). Unfortunately, other pressing priorities limited implementation of this good intention. The IPCC has indicated its review editors for the AR5 will give higher priority to communicating uncertainty in response to the IAC 2010 review of IPCC procedures. Unfortunately, they, too, are likely to be overburdened with multiple responsibilities, with the result that the treatment of uncertainty is likely to remain inconsistent and inadequate. Dedicated resources, including involvement of experts in decision and risk analysis in each writing team, will be required to make better progress.

4 A solid foundation

In soliciting this article, the editors suggested that I might reflect on the guidance’s “success, failure, or something in between”, an assignment that raises the question of the appropriate criteria for judging success. The obvious choice is to examine whether we succeeded in improving (1) characterization of uncertainty by the author teams as well as (2) communication with the end-users of the assessment.

On the first score, the TAR guidance succeeded in improving evaluation of confidence and uncertainty, for example by raising awareness of common pitfalls, reducing use of indifferent medium-confidence statements, and harmonizing the approach used in a number of chapters. Did we accomplish all that we wished and get everything “right” in the guidance document? Clearly not, but looking back, the TAR guidance was a good start and has provided a remarkably solid foundation for subsequent efforts to improve the characterization of the state of science for public policy and decisionmaking. It is possible that some of the issues that arose could have been alleviated with different approaches for calibration of confidence levels (e.g., a betting “odds” formulation or “confidence thermometers”), entirely different qualitative terms, or alternative approaches for sequencing the steps in the recommended process. It would have been helpful to give more emphasis to different dimensions of evidence that contribute to confidence (e.g., theory, observations, model results, and consensus, which structured the proposed “radar plot” graphic), as well as when to use the different approaches. But given the strong and sometimes divergent views on how to describe uncertainty in the different participating research communities, it seems unlikely that this first effort could have led to consistency of approach throughout the report, especially given the limited resources available to support implementation.

Subsequent efforts have built on the foundation of the TAR guidance, sometimes setting off in new directions based on experience in the last assessment or advances in the decision

analytic or risk communication literature, but largely remaining consistent with its approach. In the revised uncertainty guidance paper for the AR4 (IPCC 2005), the quantitative likelihood scales, confidence levels, and qualitative levels of understanding of the TAR were updated. The AR4 guidance made a more careful distinction between levels of confidence in scientific understanding and the likelihoods of occurrence of specific outcomes, solving some issues but creating others. In preparation for the AR5, the guidance was revised again, building on experience in the previous assessment rounds (Mastrandrea et al. 2010). This version in particular shows improved clarity about which approaches to use when, and includes explicit acknowledgment of the multiple dimensions of evidence that could contribute to confidence. But I have “very high confidence” (or is that “high confidence”—how would I know?) that the decision to calibrate the confidence scale *qualitatively*, not quantitatively, will create new examples of some of the well-documented confusions discussed above.

On the second objective, improving communication with end users, the challenges are formidable. Moving towards greater consistency across the elements of the reports is generally viewed as an important objective (e.g., IAC 2010), but the interviews conducted with likely users that were cited earlier indicate problems with the particular terms we recommended, and perhaps even more broadly, with the use of qualitative terminology at all, instead of simple quantitative approaches.

The ultimate test of the success of the guidance is improving how policy and public discourse incorporates the information being provided by an assessment. This objective is extremely challenging to meet because of the political potency of uncertainty. At the time that Steve and I began to lay the foundation for the guidance in 1996, the attack on the integrity of climate science in the House of Representatives, partly based on allegations of under-reported uncertainties, was in full swing and led to passage of legislation requiring ‘sound science’ as well as delays in implementing climate policy. The science community continues to be caught between different groups that have discovered that either exaggerating or minimizing the seriousness of the issue can mobilize their supporters, enhance fund raising, and win elections. Given the high stakes, it is essential to continue efforts to improve the quality of societal debate and the application of science in policy formulation. This will require assessors and users to work together.

5 Raising the bar in science for public policy and decisionmaking

The challenges for uncertainty characterization and communication are evolving, and meeting them successfully will continue to determine the utility and the credibility of climate science in the eyes of its users. For example, Nature magazine recently editorialized on the importance of providing adequate characterization of uncertainty to enable decisionmakers, who are eager for certainty in the face of high stakes and politically fraught decisions, to accurately interpret results from regional climate and impacts models (Nature 2010). Many of the authors cited in this paper and others (e.g., Oppenheimer et al. 2007; CCSP 2009) have offered suggestions for improving this aspect of assessments and meeting these broader challenges. Steve would undoubtedly have had numerous ideas as well. Drawing on his work, and that of other colleagues, I offer a number of ideas that could contribute to more accurately and openly assessing the scientific content of public policy and decisionmaking, including:

- 1) Focus on well-specified questions and decisions and provide information on the prospects for reducing uncertainty on decision-relevant time frames—assessment of uncertainty has no meaning if not considered with respect to a specific decision criterion or science question;
- 2) Include experts in decision analysis and risk communication as chapter lead authors in assessments—otherwise overall performance is unlikely to improve, given competing priorities for the time and attention of lead authors;
- 3) Take communication more seriously—test proposed approaches for conveying confidence and likelihoods before using them, refine key messages to respond to pre-existing “mental models” of users, and graphically represent uncertainty when feasible;
- 4) Use expert elicitation more widely in the IPCC and other climate assessment processes—concerns that elicitation is “subjective” and makes knowledge seem more conditional continue to impede application of this rigorous and insightful methodology;
- 5) Encourage exchange of information across research disciplines regarding approaches used to characterize uncertainty—most researchers know little about the approaches used in other fields, which will hinder synthesis and evaluation of integrated model results.

Improving the way that climate science evaluates scientific uncertainty and conveys it to policymakers and other users was one of Steve’s passions. He felt it was the key to building confidence in climate science and establishing the basis for effective management of the many risks posed by climate change. On more than one occasion, he said he considered the uncertainty guidance to be among the most significant and satisfying of his many projects. Steve established a rich and important legacy through this effort, as well as by encouraging debate and discussion of new methods in the pages of *Climatic Change*. We must build on this part of Steve’s legacy to help us improve use of scientific information in policymaking and improved risk management of an uncertain future climate.

References

- Agrawala S (1998a) Context and early origins of the intergovernmental panel on climate change. *Clim Chang* 39:605–620
- Agrawala S (1998b) Structural and process history of the intergovernmental panel on climate change. *Clim Chang* 39:621–642
- Allen M, Raper S et al (2001) Uncertainty in the IPCC’s third assessment report. *Science* 293:430–433
- Allen MR, Booth BB et al (2004) Observational constraints on future climate: distinguishing robust from model-dependent statements of uncertainty in climate forecasting. IPCC Risk and Uncertainty Workshop, Maynooth, Ireland
- Budescu DV, Broomell S et al (2009) Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychol Sci* 20:299–308
- CCSP (2009) Best practice approaches for characterizing, communicating, and incorporating scientific uncertainty in decisionmaking. A report by the climate change science program and the subcommittee on global change research. M. G. Morgan, H. Dowlatabadi, M. Henrionet al. Washington, DC, National Oceanic and Atmospheric Administration: 96
- CIA (1964) Words of estimative probability. *Studies in intelligence*. Washington, DC, Central Intelligence Agency: 12
- Dessai S, Hulme M (2004) Does climate adaptation policy need probabilities. *Climate Policy* 4:107–128

- Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy. Kluwer, Dordrecht, 229 pp
- Giles J (2002) Scientific uncertainty: when doubt is a sure thing. *Nature* 418:476–478
- Grubler A, Nakicenovic N (2001) Identifying dangers in an uncertain climate. *Nature* 412:15–15
- Ha-Duong M (2003) Imprecise probability bridges scenario-forecast gap. Pittsburgh, PA, 15 pp
- Hall J, Fu G et al (2007) Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties. *Clim Chang* 81:265–281
- Hawkins E, Sutton R (2010) The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*: 1–12
- IAC (2010) Climate change assessments: Review of the processes and procedures of the IPCC. InterAcademy Council, Amsterdam
- IPCC (2001a) Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, England
- IPCC (2001b) Climate change 2001: Impacts, adaptation, and vulnerability. Contribution of working group II to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, England
- IPCC (2001c) Climate change 2001: Mitigation. Contribution of working group III to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, England
- IPCC (2001d) Climate change 2001: Synthesis report. Contribution of working group I, II, and III to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, England
- IPCC (2005) Guidance notes for lead authors of the IPCC fourth assessment report on addressing uncertainties. Intergovernmental panel on climate change, Geneva
- IPCC (2007) Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge United Kingdom and New York, NY, USA
- Kriegler E, Held H (2005) Utilizing belief functions for the estimation of future climate change. *Int J Approx Reason* 39:185–209
- Lempert R, Schlesinger ME (2001) Climate-change strategy needs to be robust. *Nature* 412:375
- Manning MR (2003) The difficulty of communicating uncertainty. *Clim Chang* 61:9–16
- Manning MR, Petit M (2003) A concept paper for the AR4 cross cutting theme: uncertainties and risk. Geneva, Intergovernmental Panel on Climate Change: 14
- Mastrandrea MD, Field CB et al (2010) Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Intergovernmental Panel on Climate Change, Geneva
- Min S-K, Simonis D et al (2007) Probabilistic climate change predictions applying Bayesian model averaging. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365:2103–2116
- Morgan MG (1998) Uncertainty analysis in risk assessment. *Human and Ecological Risk Analysis* 4(1):25–39
- Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge; New York
- Morgan MG, Keith DW (1995) Subjective judgments by climate experts. *Environ Sci Technol* 29:468–476
- Moss RH, Schneider SH (2000) Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. Cross-cutting issues in the IPCC Third assessment report. R. Pachauri, and Taniguchi, T. Tokyo, Global Industrial and Social Progress Research Institute for IPCC, pp 33–52
- Nature (2010) Validation required. *Nature* 463:849
- NDU (1978) Climate change to the year 2000. National Defense University, Washington DC
- NIC (2007) Iran: nuclear intentions and capabilities. N. I. Council. Washington, DC: 9
- Nordhaus WD (1994) Expert opinion on climatic change. *Am Sci* 82:45–51
- Oppenheimer M, O'Neill BC et al (2007) The limits of consensus. *Science* 317:1505–1506
- Patt A (2007) Assessing model-based and conflict-based uncertainty. *Glob Environ Chang* 17:37–46
- Patt A, Dessai S (2005) Communicating uncertainty: lessons learned and suggestions for climate change assessment. *Comptes Rendus Geosciences* 337:425–441
- Patt AG, Schrag DP (2003) Using specific language to describe risk and probability. *Clim Chang* 61:17–30
- Petersen AC (2006) Simulating nature: a philosophical study of computer-simulation uncertainties and their role in climate science and policy advice. Ph.D Dissertation, Vrije Universiteit, 220 pp
- Pidgeon N, Fischhoff B (2011) The role of social and decision sciences in communicating uncertain climate risks. *Nature Clim Change* 1:35–41
- Pittock AB, Jones RN et al (2001) Probabilities will help us plan for climate change. *Nature* 413:249–249

- Politi MC, Han PKJ, et al (2007) Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making* (Sep-Oct):681–695
- Reilly J, Stone PH et al (2001) Uncertainty and climate change assessments. *Science* 293:430–433
- Risbey J, Kandlikar M (2007) Expressions of likelihood and confidence in the IPCC uncertainty assessment process. *Clim Chang* 85:19–31
- Schneider SH (2001) What is 'dangerous' climate change? *Nature* 411:17–19
- Swart R, Bernstein L et al (2009) Agreeing to disagree: uncertainty management in assessing climate change, impacts and responses by the IPCC. *Clim Chang* 92:1–29
- Titus JG, Narayanan VK (1995) Probability of sea level rise. Environmental Protection Agency, Washington DC, 186 pp
- USGCRP (2000) US national assessment of the potential consequences of climate variability and change. US Global Change Research Program, Washington DC