Elements of Change 1996

AGCI Session II: Characterizing and Communicating Scientific Uncertainty

Session Chairs: Dr. Richard H. Moss and Dr. Stephen H. Schneider

July 31 to August 8, 1996

Session Synthesis Essay: Characterizing and Communicating Scientific Uncertainty: Building on the IPCC Second Assessment

Richard Moss and Stephen Schneider, Session Chairs

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 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?

The Case of the IPCC Second Assessment Report

Interest in climate change, potential impacts, and adaptation /mitigation policy options increased dramatically during 1995-96. While there are many explanations for this, one contributing factor was the conclusion, reached by the Intergovernmental Panel on Climate Change (IPCC) in its Second Assessment Report (SAR)(IPCC 1996a-c) that, even considering remaining uncertainties, "the balance of evidence suggests that there is a discernible human influence on global climate." This conclusion, which was negotiated over a period of a year by hundreds of scientists and policymakers, acknowledged the strong belief of most experts that human modification of atmospheric composition has led to noticeable climatic effects and likely significant climate change in the decades ahead. Though not especially controversial in the scientific community, the statement created a maelstrom of both support and criticism from a variety of interest groups who seemed confused by the different ways in which uncertainties and knowns were explained in the technical chapters and the Summary for Policymakers (e. g., see Science, Nature, 1996).

How can science be most useful to society when evidence is incomplete or ambiguous, the judgments of experts vary, and policymakers seek guidance and justification for courses of action that could cause significant societal changes?

The most recent IPCC report and its predecessors provided "best estimates" of possible climatic futures, as well as a broad range of plausible outcomes, including possible "outlier" events. The implications encompass consequences of climate change that range from mildly beneficial to potentially catastrophic changes for ecosystems and human activities such as water management, development of coastal margins, and agriculture. Although more confidence was naturally expressed in outcomes near the centers of those wide ranges between the high and low outliers, some interest groups understandably focused on possible extreme outcomes, which sharpened the debate and created substantial contention.

The purpose of the IPCC and other assessments of scientific research is to convey to interested publics, including decision-makers, advisors, the media, private-sector businesses, and environmental /community groups, the most up-to-date information available. One of the major challenges is that the assessments necessarily must present a snapshot of information which is continuously evolving. At the time of preparation of the SAR, the uncertainties included, for example, the possibilities of large and/or abrupt climate changes and/or technological breakthroughs that could radically reduce emissions abatement costs in the future. Given the use of the IPCC reports in policy making, and the need of decision-makers to determine their response to the risks of climate change before all uncertainties can be resolved (even in principle) to the satisfaction of every interest group, the available information, imperfect as is may be, must be synthesized and evaluated at periodic intervals.

Thus, a great deal of importance is attached to the need to assess and explicitly distinguish which aspects of technical controversies that affect our understanding of these uncertainties are well understood and enjoy strong consensual support, which aspects are somewhat understood, and which are highly speculative. Unfortunately, in the media and political debates, such degrees of certainty and uncertainty often become blurred. As a result, the nuanced characterization of uncertainty that might occur in professional assessment is often mis -translated into the appearance of scientific cacophony in the public arena.

The nuanced characterization of uncertainty that might occur in professional assessment is often mis-translated into the appearance of scientific cacophony in the public arena.

At the same time, scientists themselves struggle with the highly subjective and qualitative nature of the assessment process, preferring, by tradition, to study individual components of problems that can be tested, rather than the necessarily more difficult synthesizing of these components of relevance to decision-makers and the public. Qualitative descriptions of the level of certainty attached to a particular finding terms such as "almost certain," "probable," "likely," "possible," "unlikely," "improbable," "doubtful," "almost impossible" mean different things to different people and hence are not precise descriptors, and they are sometimes used rather inconsistently or uncritically in assessments (let alone by the general public and the media). Individuals and groups often use simplifying assumptions or heuristic procedures for making judgments about uncertainty. The consequence of this can be a tendency towards overconfidence in the likelihood of median outcomes and a tendency to underestimate the probability of outlier events or surprises.

In order to examine previous practice in characterizing uncertainties, as well as opportunities for improving the assessment process, a group of researchers and analysts met for nine days at the Aspen Global Change Institute (AGCI) in August 1996 to investigate ways of communicating uncertainty within specific content areas of climatic assessment. The group included independent scholars, statisticians, decision analysts, media and policy analysts, and a number of Lead Authors from all three working groups of the IPCC SAR, including researchers from the physical, biological and social sciences. Our overall goal was to examine possibilities for achieving greater consistency in evaluating the judgments of scientific experts and in communicating these judgments to non-specialists. Serving as Co -Chairs of the workshop, we facilitated discussions on four basic sets of questions:

1 What approaches to establishing uncertainty ranges and confidence levels were used in the preparation process for the SAR (IPCC 1996)? How did these approaches and other factors affect the conclusions and ranges presented?

2 What approaches could be used to represent the center, the body, and the range of informed technical opinion in future assessments (including quantification of uncertainties)?

3 How do uncertainty and the presentation of levels of confidence translate into media coverage and policy debates? What approaches are used in other fields or could be developed to communicate more effectively the nuances of uncertainty as understood by climate experts, impacts specialists and policy analysts, to the media, policymakers and the public at large?

4 What recommendations for improving future IPCC or other international or national assessments could be developed by authors and experts?

Our overall goal was to examine possibilities for achieving greater consistency in evaluating the judgments of scientific experts and in communicating these judgments to non-specialists.

The workshop took advantage of the progress being made in decision sciences, in which the ability to deal formally with uncertainty is improving. These formal approaches to incorporation of new information into the policy process via adaptive strategies (*e. g.*, see Schlesinger, in this report) provide a promising framework that involves not only provision of new data, but also changes in model structure, validation and aggregation. The analyses and methodological alternatives and suggestions for improvements in the assessment process from the participants make up the most substantial portion of this report (see presentation summaries which follow).

This summary essay synthesizes the analyses and suggestions made during the AGCI meeting. In the next several sections we will:

1 review briefly the recent IPCC process and several other climate change assessments as examples of how formal assessments have treated uncertainty,

2 describe how estimates of central tendencies and ranges have been made in previous assessments and reports,

3 briefly review alternatives for more formal and consistent definitions of outlier uncertainties, and

4 discuss the potential applicability of these methods to future assessments, in particular the Third Assessment Report (TAR) of the IPCC, scheduled to be completed during 2000-2001.

Formal approaches to incorporation of new information into the policy process via adaptive strategies provide a promising framework that involves not only provision of new data, but also changes in model structure, validation and aggregation.

IPCC: Background and Case Studies from the SAR

The IPCC was established by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) in 1988, to assess scientific information about climate change relevant for international and national policy formulation. The IPCC produced its first full assessment report in 1990, a supplementary report in 1992, a special report on radiative forcing of climate in 1994, and its Second Assessment Report in 1995. It is also producing a series of Technical Papers and Special Reports in 1996-98.

The IPCC operates at two levels: as a formal intergovernmental body, and as a scientific and technical assessment body.

Intergovernmental body: Representatives of each government involved in the climate change negotiations meet in formal plenary sessions to approve the topics for assessment and the workplans for preparation of the reports. They also review and "accept" (as a whole, rather than on a line-by-line basis) the detailed reports, as well as "approve" (on a line-by-line basis) the Summaries for Policymakers that highlight the policy implications of the reports.

Scientific/technical assessment body: The body of scientists and technical experts involved as lead authors, contributors, and reviewers of IPCC assessments represent academia, governments, industry, and environmental organizations around the world.

The process used to prepare IPCC reports can be quite lengthy, often spanning two or more years. The process is initiated by national governments which commission the reports, either through the IPCC itself or through the subsidiary bodies to the U. N. Framework Convention on Climate Change (UNFCCC), approve preliminary outlines which set out the topics to be addressed, and nominate experts as potential lead authors, contributors or reviewers. Non-governmental organizations (NGOs) from both the environmental and business sectors also participate in this process. Lead and contributing authors are chosen for each chapter from among the nominations by the Bureau of the Working Groups responsible for overseeing the report. Authors and reviewers come from many countries and are trained in disciplines ranging from atmospheric chemistry to zoology and engineering to economics. The Bureau works to achieve substantive and geographical diversity. Insofar as practicable, the composition of a group of Lead Authors reflects a balance among different points of view, and includes experts from both industrialized and developing countries.

The IPCC operates at two levels: as a formal intergovernmental body, and as a scientific and technical assessment body.

During the preparation of reports, authors are charged with reviewing the most up-to-date scientific information, reconciling competing views where possible, and characterizing the range of perspectives and the reasons for disagreements within author teams when consensus does not exist. Lead authors are instructed by the IPCC not to advocate particular points of view, but rather to characterize the range of scientific information available; however, no formal roles (*e. g.*, mediators, evaluators, proponents, etc.) are assigned. Text is prepared by Lead Authors or solicited by the authors from other contributors; unprompted submissions are also accepted for consideration. As will be discussed below, the process followed by different writing teams to arrive at an agreed-upon text which describes areas of both agreement and disagreement varies considerably, from Working Group to Working Group, and issue to issue.

The IPCC documents typically undergo two stages of peer review, first by experts and subsequently by governments and accredited organizations. The expert review is open to those who have established research or technical credentials in a field related to the chapters of the report being reviewed. Effort is made to provide opportunities for input from scientific skeptics. The documents are made freely available to the full range of stakeholder groups, including environmental and industry NGOs. The Working Group Bureaus have been responsible for ensuring that comments received from reviewers are considered by the authors (a proposal under consideration for the next assessment would give this responsibility to editorial boards). Authors are requested to document their responses to comments, and these are kept on file in the relevant Working Group Support Units. The government review is open to all participating governments, as well as accredited NGOs and experts who participated in the first round of review.

Authors are charged with reviewing the most up-to-date scientific information, reconciling competing views where possible, and characterizing the range of perspectives and the reasons for disagreements.

The final stage of the report process involves review of the report by an intergovernmental meeting of the Working Group and/or IPCC plenary, where representatives speak for their governments or organizations. Most chapter authors usually attend these sessions as well, to interact with the government representatives over the report. The plenaries are responsible for ensuring that the full reports have been prepared according to established IPCC procedures that ensure openness and balance. They are also responsible for "approving" on a line-by-line basis the report's Summary for Policymakers, which is constructed by consensus in a manner that will make it consistent with the underlying technical report, representing a balance among viewpoints or interpretations of the science.

Making Collective Judgments

Collective judgments about the degrees of certainty associated with particular conclusions have been an important part of the IPCC assessment process. This is because there are no precise, mathematical approaches for simply combining the varying conclusions that different experts have reached regarding such exceedingly complex issues as the sensitivity of climate to increases in greenhouse gas (GHG) concentrations, the amount by which observed temperatures have already changed at a global mean level, or the reductions to per capita GDP that might result from climate changes associated with doubled concentrations of GHGs. Collective judgments were reached in a variety of ways, depending on the subject domain and the particular set of lead authors. The following examples from each of the three IPCC working groups Working Group I (The Climate System), Working Group II (Impacts, Adaptation, and Mitigation) and Working Group III (Social and Economic Dimensions of Climate Change) (IPCC 1996 a, b and c, respectively) illustrate the diversity of approaches used.

The fact that there is variation in approach across issues and chapter writing teams does not mean that the judgments provided in the assessment are irreproducible, of poor quality, or misleading IPCC authors developed qualitatively reasonable ranges and mean values through open, iterative discussions. Given similar information, these or very similar estimates would likely have been reached by different teams of Lead Authors acting independently. Here we demonstrate that improvements to the assessment process are possible, based in part on recent advances in statistics and methods for subjectively estimating probabilities of different events or conclusions. In our view, these improvements could be usefully applied to estimation of particularly important parameters or judgments regarding key issues. The techniques could be especially useful to improve the clarity and consistency of the estimates of the outliers or end points of the ranges of estimates generated by the writing teams. We also argue that further research into additional methods for developing collective judgments of certainty is needed in order to expand the range of approaches available to scientific assessors.

IPCC authors developed qualitatively reasonable ranges and mean values through open, iterative discussions. Given similar information, these or very similar estimates would likely have been reached by different teams of Lead Authors acting independently.

Case 1: "Climate Sensitivity"

For almost 20 years, a significant reference point in scientific assessments of climate change has been the estimate of "climate sensitivity" the globally averaged surface temperature response that eventually can be expected to occur (*i. e.*, in equilibrium) if CO 2 were to double from pre-industrial levels and remain at this concentration indefinitely. This factor, "T2x," was first identified in a 1979 National Research Council report (NRC 1979) as 1.5-4.5°C warming (for a discussion of how this value was determined, see

Schneider, in this report). This estimated range for climate sensitivity has remained essentially invariant through various assessments over the years, not because the studies that have taken place every few years revealed the same exact quantitative judgments, but largely because the relative changes from assessment to assessment were too small for scientists to justify changing this range on any plausible scientific grounds. While some changes in the underlying science have occurred (*e. g.*, new formulations of cloud or biophysical parameterizations), small (±0.5°C) changes to the outliers (end points) of the ranges have seemed unjustified even frivolous to many assessors, in view of the absence of fundamental new data or tested theory. Thus, the range has remained the same.

Improvements to the assessment process are possible, based in part on recent advances in statistics and methods for subjectively estimating probabilities of different events or conclusions.

Of course, subjectivity is inherent in complex systems analysis, but the process of achieving more consistent aggregate scientific judgments is critical to establishing more meaningful and credible ranges of potential outcomes like T2x. More consistent estimates of the endpoints of a range for any variable would minimize misunderstandings of the intent of the Lead Authors of assessments such as IPCC 1996a, and would, in turn, reduce the likelihood that interest groups could misunderstand or occasionally, misrepresent the scientific assessment process. However, such attempts to achieve more consistency in evaluating the subjective judgments of appropriate experts in specifying outlier values (or the distribution of subjective probabilities within and beyond the outlier estimates) have not received as much attention as we believe they deserve. Assessors should be encouraged to state whether the range of estimates provided for T 2x (or other parameters) is within some percentile limits (*e. g.*, 10 and 90 percent) or within one, two or three standard deviations of the mean (and, if possible, to provide the shape of the subjective probability distribution). This would make the ranges more meaningful to other scientists, the policy community and the public, and would help to increase the likelihood that the assessors themselves all have had the same level of probability in mind for the estimated range.

Case 2: Observed Changes in Surface Air Temperature

Another example from Working Group I is the estimate of 0.3° - 0.6°C surface air temperature warming for the world since the late 19th Century. This range was maintained in the SAR (1996a) assessment at the same value as in the 1990 first assessment, even though the years 1990, 1991 and 1995 were among the highest in the record. The AGCI Workshop participants familiar with the process suggested that it was more a matter of strong personalities not wishing to defend to their colleagues any change than a collective vote of the knowledgeable group which led to the range remaining unchanged. It is our understanding that a number of scientists in the process may have preferred changing the warming trend estimate to 0.4°- 0.7°C, but a sense of collegiality, combined with the difficulty of scientifically justifying changes to a range which was not initially chosen on consistent criteria for picking outlier endpoints, simply resulted in the range remaining constant. This is another example where more formal analytic methods may have led to a slightly different outcome.

Attempts to achieve more consistency in evaluating the subjective judgments of experts in specifying outlier values (or the distribution of subjective probabilities within and beyond the outlier estimates) have not received as much attention as we believe they deserve.

Case 3: Detection/Attribution

One of the major areas of advance in IPCC 1996a came in the area of detection and attribution of climate change. "Detection" is the process of demonstrating that an observed change is highly unusual in a statistical sense (analogous to measuring one's body temperature and determining that it is significantly higher than the normal). "Attribution" is the process of establishing cause and effect (analogous to explaining *why* one's body temperature is elevated once its baseline, "normal," has been established). While major scientific uncertainties still exist with regard to estimates of the magnitude, pattern and evolution of different forcings, IPCC 1996a concluded that the Earth is warming and that its mean temperature is warmer than it has been since at least the 15th Century; that observed geographical and vertical patterns of temperature change are becoming increasingly similar to model predictions that incorporate combined GHG warming and sulfate aerosol cooling effects; that the observed patterns are different enough from those due to natural variability; and that observations and model predictions generally agree in overall magnitude and timing of change.

The process used by the authors of WG I Chapter 8 on Detection of Climate Change and Attribution of Causes was statistically rigorous in fact, it was more rigorous than that used in many other chapters or most previous assessments. In addition, the key conclusions were supported by multiple lines of evidence and were not particularly surprising to most members of the informed scientific community given increases in atmospheric concentrations of GHGs. Yet the chapter's findings resulted in what one of the Lead Authors referred to as "a nightmare media aftermath" of controversy.

Why did the chapter produce such a commotion, and what, if anything, should we learn about the use of more consistent procedures for estimating levels of confidence and confidence ranges? Extracting lessons from the case is difficult due to its uniquely charged nature. Not only does detecting a human influence on climate cross a psychological/cognitive barrier (*i. e.*, humans can and are changing the planet on a global scale), but this conclusion also raises the issue of a need for potentially significant changes in economic and technological patterns that could affect many important interests. In addition, the case is clouded by perceived although not actual violations in the rules of procedure usually followed to prepare and edit chapters. The attempt to be as up-to-date as possible also meant that some of the newest materials had not been vetted for very long in the scientific literature and the community; in addition, permissible editorial changes were made following the final plenary session at which the chapter was discussed.

The key conclusions were supported by multiple lines of evidence and were not particularly surprising to most members of the informed scientific community. Yet the chapter's findings resulted in what one of the Lead Authors referred to as "a nightmare media aftermath" of controversy.

A subtle but critical distinction in the IPCC rules between "acceptance" of chapters (which does not require line-by-line approval) and "approval" of the Summaries for Policymakers (which does require line-by-line agreement to text) also contributed to the confusion. Some critics were either unaware of or chose to overlook this distinction, and criticized changes that were made to the text in the late stages of preparing the report. These changes were completely permissible, however, since the chapters are not "approved" on a line-by-line basis (a practical impossibility for a report of many hundred pages). (Edwards and Schneider, 1997, have noted that scientists typically bestow greater levels of trust in their colleagues to respond to critiques than is typically tolerated in legal or political proceedings an epistemological professional difference that contributed to political attacks on the IPCC process after the SAR was written.) Moreover, the conclusions were first communicated to the public in The New York Times , months ahead of the official release of the report, and this may have induced suspicions of procedural irregularity. This situation led to a well-orchestrated attack on the IPCC process by a number of skeptics, including some scientists, who believed (or at least stated) that their views were not adequately represented in the report. The IPCC responded vociferously to these charges, perhaps exacerbating the problem by calling more attention to the attacks than they otherwise may have received (for further discussion, see MacCracken, Edwards and Schneider, in this report).

The conclusion raises the issue of a need for potentially significant changes in economic and technological patterns that could affect many important interests.

From the point of view of lessons to be learned in shaping treatment of uncertainties and outlier ranges, this case seems to illustrate the fact that some conclusions, no matter how statistically rigorous, will produce controversy in the political arena. Clarifying procedures for preparing and revising IPCC reports so that all stakeholders are familiar with them, and developing clear procedures within the IPCC for responding to media controversies, may do just as much to avoid such controversies in the future as would employing advanced techniques for aggregating and communicating to general audiences the opinions of experts.

Case 4: Impacts of Climate Change on Vegetation and Ecosystems

The primary controversies in projections concerning the future dynamics of terrestrial ecosystems in response to climate change involve (1) the decline of terrestrial biodiversity and (2) the role of the terrestrial biosphere in the global carbon cycle. (Examinations of climate impacts on animals are in their infancy *e. g.* , Root in this report, Root and Schneider, 1993, and will be a stronger focus in future reports.) Impacts on ecosystems and the ability of various species to migrate in response to climate change have been investigated through models which attempt to determine the implications of a number of competing processes that affect the rates of change to which species can adapt, and hence to determine the sorts of ecosystems which potentially could persist in different locations. In the case of the carbon cycle, uncertainties preclude, at present, a confident assessment of the current or future role of the terrestrial biosphere as either a source or a sink of carbon. Some evidence suggests that the biosphere has been a source of CO2 emissions during natural warmings of the past 250,000 years, producing a positive climate feedback leading to further warming. Other evidence suggests that, since CO2 stimulates growth of vegetation and phytoplankton, an increased CO2 concentration will induce an increased carbon sink, providing a negative feedback on the global carbon cycle by moderating the projected growth in the CO 2 concentration. Depending on the interaction of effects in the transition from current to future climate, there is also the potential for a "pulse" of carbon to be released from dying forests (but not to be fully reabsorbed from the atmosphere in new forests for many decades or centuries), creating a positive or amplifying feedback.

From the point of view of lessons to be learned in shaping treatment of uncertainties and outlier ranges, this case seems to illustrate the fact that some conclusions, no matter how statistically rigorous, will produce controversy in the political arena.

In IPCC 1996a and b, Working Group I and II authors writing chapters covering terrestrial ecosystems contended with a number of these issues. Both groups estimated the potential effects of climate change on the carbon cycle, while Working Group II authors estimated, *inter alia*, the percentage change in vegetative cover for specified climate change scenarios. These latter estimates were based on group discussions of model results and shared agreements regarding their interpretation, rather than any formal statistical procedure for aggregating subjective probabilities (again, the lack of formal methods for objectively assessing outlier ranges limits the techniques available for application). While those involved in the discussions were familiar with the literature on both modeling and observations, the ranges ultimately chosen were not based on specified subjective probabilities for outliers. Methods to develop a judgment as to whether the outliers were within some percentile range or within one, two or three standard deviations of the mean (or to depict the shape of the subjective probability distribution) would have added clarity to the assessment.

Case 5: Impacts of Climate Change on Agriculture and Food Security

Another controversial aspect of the impact assessment concerned the adaptive potential of agriculture to climatic change and the potential implications for regional and global food security. The assessment of these issues involved a wide range of controversies, from crop/climate /insect interactions and carbon dioxide fertilization to the socioeconomic conditions under which hunger occurs and the economic modeling of agricultural trade. IPCC 1996b reached the conclusion that "global agricultural production can be maintained relative to baseline production under climate change as expressed by general circulation models (GCMs) under doubled CO2 equilibrium climate scenarios." However, in reaching this conclusion, the authors disagreed over whether net global effects really had meaning, or were in fact meaningless and should not be reported. There were also concerns regarding both over- and under-estimation of damages in this level of aggregation. The final wording reflected a careful if imprecise compromise.

The authors disagreed over whether net global effects really had meaning, or were in fact meaningless and should not be reported. There were also concerns regarding both over- and under-estimation of damages in this level of aggregation.

Estimates of the impact of equilibrium doubled-CO 2 climate conditions were made using crop-climate models, and the assessment indicated a wide range of yield changes, from large and positive to large and negative, when compared to results under current climate conditions. Some of the variation in results is accounted for by differing assumptions in different studies regarding the extent to which farmers would be able to adapt their crops or farming techniques to the new climate, thereby offsetting some of the damages or even, in some cases, taking advantage of opportunities created by expected alterations in weather conditions. A significant debate exists in the impacts community over the rate, costs, and levels of adaptation that are possible. Furthermore, assuming that some adaptation would eventually occur, others have questioned whether it would be delayed by decades because of the high natural weather variability that would mask any anthropogenic climate signal (e. g., Schneider, 1996, Morgan and Dowlatabadi, 1996).

Another debate occurred over how to operationalize one of the key concepts employed in Working Group II "vulnerability." Vulnerability could be defined for different aspects of agriculture: yield, farmer income (affecting regional, national or global economic vulnerability), or broader issues of food availability and security. While the writing team developed a precise definition of this term based on the probability density function for climate change and a damage function that relates impacts to varying levels of climate change in practice, specific climate model results played little role in the determination of vulnerability. These determinations were based on other factors, such as identifying those populations judged as being vulnerable to hunger or famine, or identifying potential thresholds for different crops.

Another difficulty was incorporating the potential for surprises that are possible and could be important (*e. g.*, spread of a plant disease to new regions or crops). A "linear" model of food supply (which assumes that supply will grow to meet demand) may have provided a reasonable gross approximation for the recent past, but that does not mean that future assessments should not systematically consider other possibilities. As in other areas, an important next step in improving assessments would be to reflect the potential for such surprises in outlier ranges.

A significant debate exists in the impacts community over the rate, costs, and levels of adaptation that are possible.

Case 6: Aggregate Economic Impacts of Climate Change and Emissions Abatement

Of all the cases examined, conducting a quantitative assessment of the economic impacts of climate change presented the most challenges (see Chapter 6, The Social Costs of Climate Change, IPCC 1996c). The IPCC writing team was faced with an overwhelming amount of information on the potential for a diverse set of impacts in many sectors and regions, and there was no robust methodology for assigning values to each impact, aggregating them, and presenting the sum in monetary terms or as a percentage of GDP.

Valuation was difficult enough for market sectors such as agriculture or human infrastructure in coastal zones, for which markets help to establish monetary values; it was an especially serious challenge when it came to valuing impacts on non-market sectors such as ecosystems and the goods and services that they provide. Due to the absence of such a methodology, the chapter authors restricted their work to a review of the existing literature on the topic, without much discussion of alternative assumptions. Valuation was also extremely controversial when applied to the question of valuation of a human statistical life. Some argued that this value was higher in developed than developing countries, while others refused to accept this result on moral grounds; more fundamentally, some objected to even attempting to assign a monetary value to a human life. However, in spite of these and additional uncertainties associated with estimates of economic impacts, there is great pressure from the policy community to provide such estimates because

(1) they deliver a single indicator of climate change impact, and (2) they make impacts directly comparable to the costs of emission control (for further discussion, see Tol in this report, and Fankhauser, Tol and Pearce, 1997).

The team was faced with an overwhelming amount of information on the potential for impacts in many sectors and regions, and there was no robust methodology for assigning values to each impact, aggregating them, and presenting the sum.

Working Group III estimated damage costs resulting from climate change impacts to range from $5-$125 per ton of carbon emitted, the range stemming from differences in estimation techniques as well as different assumptions about the appropriate "discount rate" to use in assessing future impacts in current monetary terms (note that the choice of discount rate is a source of significant dispute, but not a source of uncertainty, because it is a political/ethical choice, and the effects of different discount rates on the estimates of climate impacts can be calculated with precision). The assessment presented a range of best guess estimates of aggregate damages. Globally, annual impacts under doubled CO2 equilibrium climate were estimated to range from 1-2 percent of GDP; regionally, the annual impacts on GDP were estimated to range from slightly positive (for the former USSR and perhaps China), to mildly negative (for OECD countries), to as much as negative 10 percent (for Africa).

It is extremely important to note that ranges associated with these impacts estimates simply represented the range of best guesses of the authors, not their estimates of the full range of potential damages from low to high. No estimation of uncertainties with regard to the full social costs of climate change was made, and in fact, no systematic calculations are available in the literature as yet, as far as we know; but generally, the uncertainties are known to be large and difficult to quantify formally because of such issues as valuing non-market impacts in monetary terms (*e. g.*, Nordhaus, 1994). The lack of estimating techniques made the existing uncertainties difficult to communicate. In the opinion of one lead author, Chapter 6 of WG III conveys a message that knowledge is better developed than it, in fact, is, and that uncertainties are smaller than they actually are. It does seem likely that the ranges would not be nearly as small as portrayed in the chapter had procedures to tap outlier opinions been more formal, thus allowing lower probability minimum and higher maximum range outliers to be included in the estimates provided.

Globally, annual impacts under 2xCO_2 climate were estimated at 1-2 percent of GDP; regionally, the annual impacts on GDP were estimated to range from

slightly positive, to mildly negative, to as much as negative 10 percent (for Africa).

One of the important estimates to emerge from Working Group III is a figure (Figure 2.1) that contains a relatively wide range of climate damage estimates as a function of degree of carbon dioxide abatement. Plotted on the same axes is a narrower range of estimates of the costs to the world economy of emissions abatement (*e. g.* , carbon taxes). With regard to the relatively tight uncertainty range that Working Group III experts offered for costs of mitigation, one of the criticisms of economic analyses of such costs (e. g., Grubb et al., 1994, Goulder and Schneider 1997) is that the range of uncertainty associated with these estimates should probably be broader, reflecting uncertainties in the studies, including assumptions about the efficiency of the baseline scenario, rates of economic growth, the existence of market failures which may limit diffusion of efficient or low-emissions technology, trade patterns, energy prices, and resource availability. Another factor which has not received adequate attention is the issue of induced technological change (ITC). That is, increased prices of conventional energy would stimulate research, development and deployment of alternative energy technologies, as well as improved efficiency of all energy use, which would over time reduce costs of achieving certain abatement targets relative to those calculated in the absence of such induced technological change (see discussion in Schneider, 1997). Had such issues and uncertainties been more firmly debated and more formal procedures applied to estimate outliers, it is likely that the range given for mitigation costs would have been wider.

The range of uncertainty associated with these estimates should probably be broader, reflecting uncertainties in the studies.

Application of Techniques for More Consistent Quantitative Subjective Estimation of Outliers

In the above brief set of examples from the three IPCC Working Groups, the outlier ranges which are used to characterize the level of uncertainty associated with a particular estimate often imply potential risks or benefits with costs or values very different from best guess or central tendency estimates. Thus the values assigned to outlier outcomes are important to policymakers in that they can be taken as an indicator of the sorts of low-probability but potentially high consequence risks that are associated with different outcomes (the basis for most insurance activities). The examples outlined in Cases 1 -6 indicate that, while IPCC ranges are qualitatively reasonable and have been developed through open discussions of authors and reviewers, more attention needs to be paid to the process of developing more consistent and methodical evaluations of outlier outcomes with an explicit statement of the subjective nature of the probabilities attached. We next turn our attention to the question of what approaches might be available to develop more consistent and clear estimates of ranges and levels of certainty, and to what extent any of these techniques would be potentially applicable to the IPCC process.

While IPCC ranges are qualitatively reasonable and have been developed through open discussions of authors and reviewers, more attention needs to be paid to the process of developing more consistent and methodical evaluations of outlier outcomes.

Eliciting Expert Opinion

There is a rich literature in the decision-analytic and social -psychological disciplines of methods for eliciting subjective opinions of expert groups. To gain some insight into their potential usefulness in the IPCC, we will provide a few examples from the area of climate analysis and comment on their strengths and weaknesses. To our knowledge, there are four formal elicitations published in the area of climate change, and one of us (SHS) has been a participant in all four (NDU, 1978; Morgan and Keith 1995; Nordhaus, 1994; Titus and Narayanan, 1995).

The National Defense University's elicitation

The first took place in 1978, when the National Defense University undertook a survey requesting various scientists to present their opinions as to whether global temperatures would increase or decrease. Given the controversy at the time, the heterogeneous backgrounds of the experts selected and the ambiguity of the questions asked, it is not surprising that the opinions varied widely. The NDU study authors averaged the opinions from the participating experts, some of which leaned toward warming and some toward cooling, producing very low aggregate estimates of future climate change projections a point that was criticized later (Schneider, 1985, and Stewart and Glantz, 1985). For example, Schneider (1985) argued that it is the standard deviation or spread of the experts' opinions that has most policy relevance, since that provides a subjective metric of how different the climate of the near future could be from the current (based on the premise that difference from present is likely to cause impacts). Moreover, he questioned whether all experts surveyed were comparably knowledgeable about the underlying science of climate projections. This leads to the very controversial (and still relevant) question of devising techniques which are appropriate for aggregation of opinions from experts with heterogeneous skills, a matter we discuss below.

The NDU study authors averaged the opinions from the participating experts,

some of which leaned toward warming and some toward cooling, producing very low aggregate estimates of future climate change projections.

The Morgan-Keith study

The second study, carried out by Granger Morgan and David Keith, from Carnegie-Mellon University (CMU), was very different from the NDU study. In this case, the authors first visited each of the respondent scientists, provided in advance a background paper on climate change to each scientist, and requested comments and corrections from those scientists on the concepts of that paper. Morgan and Keith then revised their questionnaire to be sure that the content about which expert opinion was elicited would be meaningful scientifically. The authors then personally visited each of the respondents, asking each one to draw cumulative density functions displaying their subjective probabilities for a number of outcomes in the global warming debate. We concentrate here on the respondents' estimates for the climate sensitivity parameter, T2x (see Figure 2.2). These tentative initial cumulative probability functions were then discussed by each respondent and the visiting study authors, with the latter pointing out to the former precisely what the hand-drawn sketches translated into: that is, the respondent implied an x percent probability that climate sensitivity would be less than 1°C, y percent probability it was less than 4°C, and z percent probability it was less than 10°C. This feedback helped to ensure that each scientist's sketch in fact conformed with his own internal beliefs. Not only did this feedback help respondents to represent their own thinking more consistently, but was the opening exercise of a long interview.

Later in the interview, the CMU team asked each scientist to specify the components of the climate problem that they believe created the climate sensitivity and led to its uncertainty distribution. Problems such as cloud feedbacks, ocean fluxes, etc., were addressed by respondents. The study authors then checked for consistency between the range of uncertainty elicited from the scientists on the outcome (*i. e.*, T2x on Figure 2.2), and the degree of uncertainty associated with processes such as cloud feedbacks. This helped the CMU authors determine the relative internal consistency of each scientist by comparing their reasons for uncertainty in the outcome and their actual subjective probability functions as displayed in Figure 2.2. These results were then returned to the scientists, who, witnessing their own possible inconsistencies, had the opportunity, if they wished, to redraw their cumulative probability functions on outcomes such as climate sensitivity. Through this interactive process, scientists were able to rethink, reexamine and arrive at a refined estimate of their own true subjective opinions of probability in formal terms.

Through this interactive process, scientists were able to rethink, reexamine and arrive at a refined estimate of their own true subjective opinions of probability in formal terms.

While it is true that the elicitation process took four hours of interview time on each occasion, and subsequent revisits to the task, the product of this study is a remarkably consistent set of cumulative probability functions (see Figure 2.2). Interestingly, although most scientists interviewed had largely overlapping ranges of subjective probability, with means and medians for T2x close to the conventional wisdom of 2-3°C (also the best guess of the IPCC) nearly all sixteen analysts also allowed somewhere between a 5 and 20 percent chance of small climate change (less than 1°C). Moreover, most of the sixteen allowed a significant (greater than 10 percent) chance of large climate change (greater than 4°C). The two columns on the right of Figure 2.2 show the mean and standard deviation of the cumulative density functions of each scientist.

One scientist, number 5 on Figure 2.2, had a radically different distribution from the rest of the group. Scientist 5 essentially viewed climate sensitivity via a different paradigm than the others. Granger Morgan has argued that it is inappropriate to aggregate the opinions of experts into a group mean and standard deviation when experts appear to hold to different paradigms. For example, had half the experts agreed with Scientist 5 that climate sensitivity is negligible, while the other half had estimates that lay in the center of the standard range, aggregating the two would create a bimodal distribution. This distribution would have peaks around 0.5°C and 2.5°C, but its mean would be about 1.5°C. Because the mean of a bimodal distribution will often be a value that is itself unlikely, the net effect of aggregating in this case would place the group average sensitivity in between the two paradigms where neither group believes the best guess belongs. Under such conditions, Morgan and Henryon (1990) have argued that it is best to simply exhibit the two groups of elicitations in separate blocks. There is, therefore, no group average given to accompany Figure 2.2. Given the summary nature by which aggregate distributions in that study were presented, aggregating the distribution would have been misleading to some readers. Aggregating may be appropriate in other studies where readers are unlikely to be misled, such as studies that present probability density functions where the bimodal nature of the results would be evident, or Monte Carlo modeling assessments where the model employed can make efficient use of the input information even if it includes bimodal distributions.

The Titus-Narayanan EPA sea level rise study

The third study in the case of global climate change was undertaken at the Environmental Protection Agency (EPA) by Jim Titus and Vijay Narayanan (1995). This study was explicitly designed to improve upon the existing IPCC integrated assessment framework. The authors started with the set of models that IPCC uses to project emissions, concentrations, transient temperature change, and sea level rise resulting from thermal expansion and glacial melting. IPCC's projections require different groups of experts to specify low, medium, and high assumptions for various key processes (a wide variety of processes involving many different types of expertise) and the end results are the official IPCC low, best estimate, and high projections of future temperatures and sea level. Titus and Narayanan instead sought to estimate a probability distribution for temperature and sea level rise by soliciting different groups of experts to specify probability distributions for key coefficients and the likelihood of particular models.

Because the mean of a bimodal distribution will often be a value that is itself unlikely, the net effect of aggregating in this case would place the group average sensitivity in between the two paradigms where neither group believes the best guess belongs.

Like the NDU, CMU, and Nordhaus (1994) studies, Titus and Narayanan also included an effort to elicit subjective cumulative probability distribution functions of subjective probability from various experts, but in this case the focus was on obtaining probability distributions of model parameters. Only one of the parameters dealt with outcomes (*i. e.*, T2x) while the others dealt with values of partly known (but critical) parameters used in climate systems models (*e. g.*, mixing rates between upper and deep oceans, polar amplification of warming, etc.), rather than climate change outcomes. (The distinction between outcome and parameter in a model is relative. For example, the T2x parameter is the outcome of an equilibrium run of a 3-D GCM, but it is a parameter in the 1-D ocean models used by IPCC and EPA to project temperatures and sea level.)

This study not only solicited distributions for parameter ranges from climate modelers, but also treated all aspects of the problem with distribution functions. Thus, the authors began with a subjective statistical distribution of carbon emissions, fed those emissions into a carbon cycle model, fed those results into the climate modeling section just described, then fed those results into ocean mixing models, and fed those results into glacier melt models to calculate eventual probabilities of sea level rise. At each step, subjective probability distributions for key parameters were fed into the models; in some cases, reviewers also attributed probabilities to alternative models. For example, while the IPCC assessment did not include any models that allowed for a disintegration of the West Antarctic Ice Sheet, the EPA authors included some models that did so. In a few cases, reviewers even suggested models that had not been within the original model framework. Monte Carlo techniques were used to generate final probability distributions.

Titus and Narayanan sought to estimate a probability distribution for temperature and sea level rise by soliciting different groups of experts to specify probability distributions for key coefficients and the likelihood of particular models.

Titus and Narayanan only sought opinions on processes for which individual scientists were recognized experts, and allowed anyone to opt out of specifying a probability distribution for any particular parameter where the scientist lacked expertise. The reviewers fell broadly into three groups: climate and ocean modeling, polar amplification, and glacial modeling, but not every expert expressed an opinion on every parameter related to his or her field. The rationale for this approach was (1) not to ask reviewers to estimate parameters on which they have little expertise, and (2) this approach is consistent with the IPCC assessment, which had different committees responsible for emissions, concentrations, ocean modeling, and glacial processes. The questions posed to the experts concentrated on estimating the values and uncertainties in the process parameters, letting the models convert these uncertainties into outcome probability distributions for global temperatures and sea level rise.

Like the Morgan and Keith study, there was one scientist whose estimate of climate sensitivity was an order of magnitude less than the eight other scientists involved in the EPA climate model elicitation. There was also one glaciologist whose suggested parameters and model probabilities implied a 5 percent chance of a major Antarctic contribution to sea level rise while the other glaciologists provided parameters suggesting the probability to be only 0.5 percent. In both cases, the EPA authors weighted all of the scientists' opinions equally in calculating the combined distribution, to ensure that the one "contrarian" scientist's response, which resulted from his fundamentally different view of climate sensitivity, was given full effect. These aggregations were criticized in editorial comments by Keith, 1996, on the grounds that this is an inappropriate procedure when an outlier scientist clearly adheres to a different paradigm than the other scientists interviewed. Titus and Narayanan considered arguments against this approach, but, they argue, to simply ignore, for example, the outlier glaciologist's opinion because it is a minority view would be to disregard the possible risk of a large rise in sea level from rapid Antarctic disintegration. They also noted that, unlike solicitations that focus on outcomes, a modeling study in which different experts estimate different parameters must devise a means of combining opinions, and that Keith's criticism applied to all aggregation procedures. (For further discussion, see Titus, in this report.)

Despite some debate over methodologies, this pioneering study did produce a wide distribution of probabilities for the outcome of concern (sea level rise), ranging from a small probability of no rise, to a small probability of greater than a meter rise by the end of the next century. The overall shape looks somewhat like a normal distribution except it is skewed to the right allowing a small chance of a very large rise in sea level, and, as such, it encompasses a somewhat wider range than that resulting from the outlier ranges chosen by much less formal means by IPCC 1996a.

The EPA authors weighted all of the scientists' opinions equally in calculating the combined distribution, to ensure that the one "contrarian" scientist's response, which resulted from his fundamentally different view of climate sensitivity, was given full effect.

The Nordhaus study

In order to produce integrated assessments or cost-benefit analyses, attempts are made to weigh the impacts of anthropogenic climate change on environment and society (*i. e* ., climate damage) against the costs of mitigation activity, such as carbon taxes. In order to perform such optimal cost-benefit analyses, scientists need probability distributions of both costs and benefits. Nordhaus (1992) provided such a calculation based upon an assumed cost to the world economy of about 1 percent of GDP from a CO2 doubling. This climate damage estimate was criticized as arbitrary and too small by a number of scientists, which led Nordhaus (1994, the fourth study considered here) to perform an elicitation of the opinions of 19 economists and other scientists regarding the damage to the world economy from several scenarios of climate change: (a) a warming of 3°C by 2090, (b) a warming of 6°C by 2175, and (c) a more rapid warming of 6°C by 2090. The probability distribution elicited for each scenario required the respondents to provide 10 percent probability, median, and 90 percent probability estimates for the likely damage to the economy from these climate scenarios. Figure 2.3 shows this result for the 6°C warming by 2090. Nordhaus performed his elicitation by a combination of formal written questionnaires and follow-up telephone interviews. Like Morgan and Keith, he gave authors a chance to reconsider their own judgments based upon the elicitations from fellow respondents.

Nordhaus performed his elicitation by formal written questionnaires and follow-up telephone interviews. Like Morgan and Keith, he gave authors a chance to reconsider their own judgments based upon the elicitations from fellow respondents.

Two features are apparent in Nordhaus' results. First, in addition to very widespread ranges, in general, there are clusters near the ends of the range, at low and high damage levels. Nordhaus classified these according to the profession of the respondents, concluding that mainstream economists are a factor of twenty less worried than natural scientists, even about a 6°C warming in a century. (At a meeting where the results were discussed, Nordhaus quipped that those who know the most about the economy are least worried, to which Schneider counter-quipped that those who know the most about nature are most worried (*e. g.*, Schneider, 1997b, chapter 6)). Differences in training, world views, and the relative amount of damages attributed to non -market sectors (*e. g.*, biodiversity loss) contributed to these paradigm differences among groups of different professionals.

Nordhaus concluded that mainstream economists are a factor of twenty less worried than natural scientists, even about a 6°C warming in a century.

Interestingly, despite the order of magnitude difference in their absolute GDP damage estimates, which correlated with disciplinary backgrounds, both natural scientists and economists drew non-linear damage functions. That is, both felt that the 6°C warming to 2090 AD would have damages more than twice the 3°C warming to 2090 case. It is interesting to contrast this with an elicitation by Berk and Schulman, 1995, of citizens in California's San Fernando Valley, who were asked how much they were willing to pay to avoid climate changes that were 5, 10, 15 or 20°F warmer or colder than the summertime averages in that location. These citizens were willing to pay more for larger deviations from the mean, but achieved a diminishing return in their willingness to pay, relative to the elicitation of the experts. Few said they would be willing to pay as much as $500 to prevent a climate like that of Death Valley (mean temperatures of 120°F) from occurring in the Los Angeles Basin! This is in sharp distinction to the non-linear (more like temperature difference squared; see Roughgarden and Schneider, 1997) views of the experts, whether economists or natural scientists. Moreover, Roughgarden and Schneider's analysis of Nordhaus' survey data reveals that there was a positive correlation between respondents who believed there would be large climate damages and the proportion of damages assigned to non-market sectors. Clearly, such formal techniques are useful for revealing different categories of subjective probability among various groups, but it remains unclear how to develop a widely accepted conclusion.

The AGCI group also considered approaches not yet tried in the climate-change arena, including an additional kind of formal procedure for evaluating and aggregating subjective expert opinion, based on experience with earthquake hazard assessments.

Elicitation of expert judgments from other fields

The AGCI summer session also considered approaches not yet tried in the climate-change arena, including an additional kind of formal procedure for evaluating and aggregating subjective expert opinion, based on experience with earthquake hazard assessments. Probabilistic seismic hazard analysis is a methodology for estimating the likelihood that various levels of earthquake-caused ground motions will be exceeded at a given location in a given future time period. Due to large uncertainties in the geosciences data and in their modeling, multiple models and interpretations are the norm, leading to disagreements among the experts. A process of aggregation of expert opinions, both social and scientific, was designed by the Senior Seismic Hazard Analysis Committee (SSHAC) funded by the Department of Energy, the

Nuclear Regulatory Commission and the Electric Power Research Institute and chaired by Robert Budnitz. It has been reviewed by a National Research Council committee (NRC, 1997). The objective is to obtain a single mean estimate of the annual frequency of exceedance versus peak ground acceleration at a given site, and composite uncertainty bands reflecting the informed scientific community's current uncertainty in that estimate. The process represents an alternative way of dealing with the problem of aggregating expert opinion in a particular scientific area for purposes of policy making. It has been applied recently to other subjects such as volcanic risks. (This process is discussed by Cornell, in this report.)

The first stage of the SSHAC process involves discussion by a large group of experts that reviews and weights different approaches to modeling, dropping some approaches as demonstrably wrong or as not worth pursuing further in depth and promoting others. The second stage involves the careful selection of experts for the remainder of the process who are asked to wear different hats, some as proponents of particular models or approaches and others as evaluators, who play the major role in the process by evaluating the form of uncertainty distributions on parameter values and/or alternative models. The ultimate responsibility for the success of this process lies with the Technical Facilitator/Integrator (TFI). The TFI is a single entity, sometimes an individual, but preferably a small team that is responsible for identifying the key issues and components of the analysis, structuring and directing the interaction and debates among the proponents and evaluators, conducting any necessary numerical analyses, and documenting the process followed and the results obtained.

The objective of the process is to allow researchers to be in a position to give an aggregate, composite range of the group's uncertainty.

The SSHAC process provides for clarification of objectives, consensus, and the roles of experts. It also involves numerical or mechanical aggregation schemes, and social integration procedures. The objective of the process is to allow researchers to be in a position to give an aggregate, composite range of the group's uncertainty. The process involves very strong interaction, with the TFI like a panel of court judges (see original proposal for a science court by Kantrowitz, 1967 and 1976) holding the final responsibility for defending the result as an accurate characterization of the group's collective judgment. The cost of conducting this process well for any given issue may be up to $1 million.

Despite the cost obstacle, some argued that an independent panel might be more credible in the policy community than experts who might be perceived as having vested interests in any aspect of their respective elicitations. On the other hand, AGCI participants recognized that such vested experts may be uniquely qualified to estimate the subjective probabilities of outcomes they specialize in studying relative to more "neutral" judges who are less vested, but also less knowledgeable. Of course, large committees of experts, acting as both assessment witnesses and peer reviewers, can minimize the likelihood of vested interest bias progressing very deeply into the aggregated assessments. This is another area in which different degrees of peer trust among professional communities (*i. e.*, scientific versus legal versus political) may strongly influence what each community believes to be appropriate assessment process design (see discussion in Edwards and Schneider, 1997).

Vested experts may be uniquely qualified to estimate the subjective probabilities of outcomes they specialize in studying relative to more "neutral" judges who are less vested, but also less knowledgeable.

Options for Improvements

We considered three basic steps that assessment bodies such as the IPCC should consider taking in order to achieve more consistent and clear subjective opinions of participating experts.

1 Improving consistency in estimating ranges and levels of confidence Most straight forward would be the presentation of consistency standards that are qualitative, but could nonetheless provide more systematic definitions of levels of confidence based on degrees of evidence and consensus (Table 2.4, Moss 1996 and Table 2.5, Trenberth 1996, see also National Academy of Sciences report, 1987).

At a minimum, employing such consistency tables would force participants to think more carefully and consistently about their subjective probabilities, and help to translate words like high, medium, and low confidence into reasonably comparable probability estimates. This step would be relatively straightforward to implement, and could improve the consistency of the subjective estimates in future assessments.

Examining these different estimates of uncertainty by each individual could help to sharpen the intuitive scientific judgments and not only add consistency, but enhance the quality and reliability of the range limits provided by assessors.

---

2 Elicitation of subjective probability distributions The second technique, following Morgan and Keith, would frame a number of explicit questions in which ranges of estimates of outcomes or parameters would be called for from authors or other participants. Formal one-on-one interviews, group interviews or mail-in questionnaires (perhaps with follow-up contacts) could be used to establish subjective probability distributions for a few key parameters or issues. In addition to soliciting outcomes (*e. g.*, estimates of climate sensitivity), processes and major uncertainties in data or theory could also be elicited (*e. g.*, Titus and Narayanan, 1996). Furthermore, questions could be designed to improve the consistency between outcome estimates and process uncertainty estimates made by each individual or group.

These probability distributions would not necessarily be published as part of the final report, but could be used in the process of preparing and revising drafts at successive author meetings. Examination of these different formal estimates of uncertainty by each individual, particularly in a group setting, could well help to sharpen the intuitive scientific judgments and not only add consistency, but enhance the quality and quantitative reliability of the range limits provided by the assessors. Not only might a single elicitation activity be undertaken, but successive meetings of the working groups could each devote a short period in which re-elicitation is attempted, based upon what each participant learned from the previous elicitation and the research that they each undertook in-between meetings, partly stimulated by the differences in each respondent's elicitation and those of his/her colleagues. Such differences are likely to cause discussions to be sharpened and focused on points of uncertainty among the participants. The privacy and formality of such procedures would also help to overcome the difficulty of dominant personalities having undue influence in small groups of people from different cultures and personalities.

3 Assessment Court Finally, an assessment court, or panel of independent judges, could be selected to listen to the debates, to review the elicitations of authors or other experts, and to provide an independent evaluation of the experts' assessment. While such procedures are more formal and cumbersome (as well as costly) than the previous two, if they resulted in higher credibility for the assessment report or more consistency in the estimation of outlier events whose damages could be nonlinearly more significant than smaller outlier ranges, then the extra effort might be worth the trouble. Even if the formal process were not applied, sensitizing authors to the various roles identified (proponents, evaluators, etc.) could help establish more systematic guidelines for participating in assessments.

What are the pros and cons of these three techniques? Under which circumstances could they be applied? No universal recommendations can be developed, as the purposes and context of each assessment varies, and techniques appropriate to one context may not be appropriate to others. However, the clarification of language, as in Tables 2.4 and 2.5, would minimize the likelihood that different participants meant different quantitative probabilities while using the same words. That strikes us as an absolute minimum requirement for future assessments.

The privacy and formality of such procedures would also help to overcome the difficulty of dominant personalities having undue influence in small groups of people from different cultures and personalities.

Formal elicitation would have several benefits. It would force people to confront their doubts and reexamine their relative uncertainties. It would provide assessors with a preliminary estimate of comparability to their colleagues thinking in quantitative terms, which, in turn, might lead to debate that could result in different (presumably better) subjective estimates than would take place in the absence of intercomparison of various authors' formal elicitations. This formal procedure certainly would help to get the opinions of more reserved participants counted equally with more extroverted ones.

The assessment court procedure would bring to the process the ostensible credibility of an independent group, eliminating the frequent criticism that the recommendations from the assessment scientists can be (even if unconsciously) motivated by self-interest, since the scientists indeed are interested parties in performing the scientific research they recommend as needed.

For the procedure to be most effective, it needs to operate in the learning mode, that is, learning by inter-comparing various authors' elicitations both at initial and follow-up meetings.

What are the drawbacks of such formal techniques? With regard to more systematic definitions of uncertainty via qualitative categories (*e. g.*, Tables 2.4 and 2.5), we cannot think of any. It seems essential, we have already argued, that this be a minimum next step. With regard to formal elicitations, a number of possible drawbacks come to mind. First, for those who do not understand the nature of subjective probability, it could convey a false sense of analytic precision, at least for a reader not recognizing that the probabilities, though quantitative, are still highly subjective. Of course this objection could be mitigated by careful writing and clear explanation of the activity.

The second difficulty is that for the procedure to be most effective, it needs to operate in the learning mode, that is, learning by inter-comparing various authors' elicitations both at initial and follow-up meetings. This puts a premium on having the same sets of authors attending most of the meetings, so that the learning process can lead to evolving and more consistent subjective probability estimates. Another potential drawback is that peer or political pressures could induce people whose elicitations might be publicly displayed to produce cumulative density functions that do not reflect their true beliefs (this objection could also be applied to the consistency table approach to assessment discussed above). However, it seems easy enough to overcome any potential problem by assigning a number to the elicitation of each respondent and displaying this number rather than the name of the scientist who offered each response. Thus, the responses would be known only to the individual scientists, and would be revealed only if a scientist chose to identify him or herself with a particular number. Another potential drawback is the (remote) possibility for political trade -offs among delegates, but once again this danger also exists in current qualitative assessments, and is best avoided by open meetings and extensive peer review practices already in place.

A difficult potential drawback is that not all authors bring the same knowledge base to bear in assessment workshops, and thus the appropriate relative weights in the aggregation of authors' opinions may not be equal. It is very divisive to try to determine whose opinion is more significant than another's, and this weighting process needed for aggregation could lead to significant internal dissension in the assessment team. However, an anonymous system of elicitation display where only the scientist knows which chart is his or hers, and vigorous discussion among participants willing to either admit their views or defend certain elicitations of others, could lead to substantial and rapid education among those less well acquainted with specific topics, thereby improving the overall quality of the final elicitations that result at the end of the iterative process. Finally, rather than averaging all results, they can be either aggregated into paradigmatically similar groups or displayed in full without averaging.

Not all authors bring the same knowledge base to bear in assessment workshops, and thus the appropriate relative weights in the aggregation of authors' opinions may not be equal.

One drawback of the assessment court approach is that a search for "neutral" experts outside of the climate community might not produce higher quality opinions, and might require that group to meet for a long time to get up to speed on the expert dimensions of the very complex sets of climate issues, thereby invalidating the extra credibility that would ostensibly derive from their alleged unbiased (even if less knowledgeable) presence. However, on politically sensitive items (such as biodiversity protection or the value of a statistical human life used in cost/benefit analyses), limited use of such independent assessment teams may, nonetheless, be worth considering.

Another major drawback of this approach is the fact that court systems in general tend to create advocates who typically ignore or denigrate data and evidence that do not necessarily support their theses. In other terms, whereas the display of expert opinions (via distributions and/or aggregations) should make use of all available information, court systems tend to create incentives for the opposing sides to either suppress or exaggerate the flaws of evidence that does not support their thesis. This may result in either an artificial level of confusion or a substantial truncation of data that lie in the middle and do not support clearly extreme positions (for more, see Paté -Cornell, 1996). Such risks can be mitigated by proper training and control by the assessment judges, but it is still not clear whether scientific assessments in general should move closer toward (or even further away from) a legal model in which opposing advocates often operate on an "it's not my job to make my opponent's case" epistemology. IPCC has encouraged its scientific assessment teams to consider the full range of evidence, not to serve as advocates.

IPCC has encouraged its scientific assessment teams to consider the full range of evidence, not to serve as advocates.

The Media and Political Context

When scientists' debates regarding uncertainties and levels of confidence in findings move into the media and political contexts, many more variables come into play. From the difficulties journalists have in understanding and communicating these issues to the public, to the use and misuse of scientific uncertainties by interest groups, an already confusing set of issues is further confounded. To use one example, some media accounts portrayed the fact that the SAR's 1995 estimate of mean annual surface temperature increase for 2100 A. D. was about 1°C lower than the temperature estimate in the 1990 report as an indication that climate was much less sensitive to GHG emissions than previously thought. However, this change in the projection was due to the inclusion of the potential effects of sulfate aerosols in the models, not to a change in the sensitivity of climate to forcing, the estimate of which remained constant (see Case 1 above). While scientists understand and explained this distinction, it is often read by much of the public to mean that the problem is not as bad as was thought a serious misunderstanding of the findings, since (a) aerosols exhibit a local and short-lived cooling effect which alters regional patterns of climate change, and (b) projections of future aerosol concentrations have a very large uncertainty.

Another example is the manner in which the detection and attribution of climate change, discussed in Case 3 above, was portrayed in the media. The IPCC chapter on this subject is one of the most statistically rigorous in the entire assessment. Its conclusions are based on multiple supporting lines of evidence. Although most of the debate within the scientific community centered on narrow concepts of statistical significance, the debate was characterized in the media and political contexts as being much broader (Edwards and Schneider, 1997). In both of these examples, much of the media portrayed what were actually debates about narrow scientific concepts or minor procedural disputes as though they involved more extensive uncertainties or were based on political advocacy.

Much of the media portrayed what were actually debates about narrow scientific concepts or minor procedural disputes as though they involved more extensive uncertainties or were based on political advocacy.

Conclusion

Assessment is not focused on original scientific results, but rather on evaluation of existing science, to address policy-related issues and to provide perspectives on the relative parts of the problem that are well understood and to distinguish these from more speculative aspects. Given the nature of the issue, climate change assessments need to encompass end-to-end coverage of a broad spectrum of issues including emissions of greenhouse gases, changes in radiative forcing, implications for global and regional climate impacts, and options for reducing emissions in energy sectors and land management. As a consequence, such assessments need to be multidisciplinary, which requires a diverse set of researchers and experts to be involved in preparing and reviewing the documents. Communication across such a broad range of disciplines about the joint probabilities of related events is a challenge; the strength of assessments can be improved by more systematic attention to the terms and approaches used in assessing and describing the degree of certainty associated with findings.

The strength of assessments can be improved by more systematic attention to the terms and approaches used in assessing and describing the degree of certainty associated with findings.

Assessments need to make estimates of the likelihood that (and time frame over which) new scientific knowledge can resolve existing uncertainties and assess the impacts of changes to physical, biological and societal conditions. They also need to assess the potential costs of those impacts relative to the potential costs of mitigation activities that would reduce the environmental changes that create the impacts (the so-called optimization process). The assessment process and its interaction with the policy community is a cyclical process over time. Since any one assessment will necessarily be incomplete in regard to all of the needed information, it is clear that rolling reassessments will be required every so often (thought to be about five year intervals by the IPCC), so that governments and other decision making groups can reassess the degree to which mitigation, adaptation, or other responses should be stepped up or delayed.

Since reassessment is needed for adaptive management, it is critical that each succeeding assessment refine the way it treats and communicates subjective estimates of uncertainty and the ranges of estimates that are provided. The proposals developed during the AGCI summer session provide useful ideas for the next step in the process of reassessing the potential implications of climate change.

Scientific uncertainty, whether about climate change policy or many other socio-technical issues, is not a problem which can be "solved." Rather, it is an ongoing challenge and an increasingly important part of the context in which policy decisions will be made. With negotiations of the Framework Convention on Climate Change progressing, it is clear that the world will not wait for virtual certainty before making policy. It is also likely that some interests will demand precisely such certainty before agreeing to implement policies that they perceive negatively affect their interests. In this context, improving the characterization and communication of uncertainty is invaluable. The AGCI sessions' contribution, in addition to specific recommendations contained in this report, also stems from improving understanding and dialog among disciplines from the natural and social sciences, the field of decision analysis, and the media and policy arenas.

Technical specialists should not assume they will be understood by nonspecialists unless considerable effort is invested in separating out what is speculative from what is likely in a clear and consistent manner. To do less is to invite confusion and contention that will delay the process of determining how societies intend to address the risks of climate change.

Scientific uncertainty, whether about climate change policy or other issues, is not a problem which can be "solved." Rather, it is an ongoing challenge and an increasingly important part of the context in which policy decisions will be made.

References

Berk, R. A., and Schulman, D., 1995, Public Perceptions of Global Warming, Climatic Change, **29**(1):1-33.

"Climate Debate Must Not Overheat," Editorial, Nature, **381**(6583): 539, 1996.

"Climate Report 'Subject to Scientific Cleansing'," Nature, **381**(6583):546, 1996.

Edwards, P. and Schneider, S. H., 1997, The 1995 IPCC Report: Broad Consensus or "Scientific Cleansing?" (Ecofables/Ecoscience , Vol. 1, No. 1, in press).

Goulder, L. and Schneider, S. H., 1997, Induced Technological Change, Crowding Out, and the Attractiveness of CO2 Emissions Abatement, Resource and Environmental Economics (submitted).

Grubb, M., Duong, M. H. and Chapuis, T., 1994, Optimizing Climate Change Abatement Responses: On Inertia and Induced Technology Development. In Integrative Assessment of Mitigation, Impacts, and Adaptation to Climate Change. N. Nakicenovic, W. D. Nordhaus, R. Richels, and F. L. Toth, eds. International Institute for Applied Systems Analysis, Laxenburg, Austria. pp. 513-534.

Fankhauser, S., R. S. J. Tol and D. W. Pearce, 1997, Extensions and Alternatives to Climate Change Damage Valuation: On the Critique of IPCC Working Group III's Damage Estimates, submitted to

Environment and Development Economics.

IPCC, 1990a: Climate Change: The IPCC Scientific Assessment . Prepared by IPCC Working Group I [J. T. Houghton, G. J. Jenkins, and J. J. Ephraums (eds.)] and WMO/UNEP. Cambridge University Press, Cambridge, UK, 365 pp.

IPCC, 1990b: Climate Change: The IPCC Impacts Assessment . Prepared by IPCC Working Group II [W. J. McG. Tegart, G. W. Sheldon, and D. C. Griffiths (eds.)] and WMO/UNEP. Australian Government Publishing Service, Canberra, Australia, 275 pp.

IPCC, 1991: Climate Change: The IPCC Response Strategies . Prepared by IPCC Working Group III and WMO/UNEP. Island Press, Covelo, California, 273 pp.

IPCC, 1992: Climate Change 1992: The Supplementary Report to The IPCC Scientific Assessment. Prepared by IPCC Working Group I [J. T. Houghton, B. A. Callander, and S. K. Varney (eds.)] and WMO /UNEP. Cambridge University Press, Cambridge, UK, 200 pp.

IPCC, 1994a: Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios. Prepared by Working Groups I and III [J. T. Houghton, L. G. Meira Filho, J. Bruce, Hoesung Lee, B. A. Callander, E. Haites, N. Harris, and K. Maskell (eds.)] and WMO/UNEP. Cambridge University Press, Cambridge, UK, 339 pp.

IPCC, 1994b: IPCC Technical Guidelines for Assessing Climate Change Impacts and Adaptations. Prepared by IPCC Working Group II [T. R. Carter, M. L. Parry, H. Harasaw, and S. Nishioka (eds.)] and WMO/UNEP. University College London, UK, and Center for Global Environmental Research, Japan, 59 pp.

IPCC, 1995: IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the IPCC, OECD, IEA, and WMO/UNEP. IPCC Working Group I TSU, Bracknell, UK, 3 volumes.

IPCC, 1996: Climate Change 1995: The Science of Climate Change . Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J. J., L. G. Meiro Filho, B. A. Callander, N. Harris, A. Kattenberg and K. Maskell (eds.)]. Cambridge University Press, Cambridge and New York, 584 pp.

IPCC, 1996: Climate Change 1995: Impacts, Adaptations, and Mitigation of Climate Change: Scientific-Technical Analyses . Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change [Watson, R. T., M. C. Zinyowera, and R. H. Moss (eds.)]. Cambridge University Press, Cambridge and New York, 880 pp.

IPCC, 1996: Climate Change 1995: Economic and Social Dimensions of Climate Change. Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change

[Bruce, J., Hoesung Lee, and E. Haites (eds.)]. Cambridge University Press, Cambridge and New York, 464 pp.

Kantrowitz, Arthur, 1967, Proposal for an Institution for Scientific Judgement, Science, **156**:763-764.

Kantrowitz, Arthur, et al., 1976, Presidential Task Force on the Science Court, The Science Court Experiment: An Interim Report, Science, **193**: 653-656.

Keith, D., 1996, When Is It Appropriate to Combine Expert Judgments? An Editorial Essay, Climatic Change, **33**(2):139-143.

Morgan, M. G., and Dowlatabadi, H., 1996, Learning from Integrated Assessment of Climate Change, Climatic Change, **34**(3-4): 337-368.

Morgan, M. Granger and David Keith, 1995, Subjective Judgments by Climate Experts, Environmental Science and Technology, **29**:468-476.

Morgan, M. Granger, and Max Henryon, with Mitchell Small, 1990, Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Policy Analysis, Cambridge University Press, New York (paperback edition 1992).

National Academy of Sciences, 1987, Current Issues in Atmospheric Change, National Academy Press, Washington, D. C., pp. 9-10.

National Defense University, 1978, Climate Change to the Year 2000 , National Defense University, Washington, D. C.

Nordhaus, W. D., 1994, Expert Opinion on Climatic Change, American Scientist, Jan.-Feb. 1994, **82**:45-51.

Nordhaus, W. D., 1992, An Optimal Transition Path for Controlling Greenhouse Gases, Science, **258**:1315-1319.

Panel on Seismic Hazard Evaluation, Committee on Seismology, National Research Council, 1997, "Review of Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts," National Academy Press, Washington D. C.

Paté-Cornell, M. E., 1996, Uncertainties in Global Climate Change Estimates. An Editorial Essay, Climatic Change, **33**(2):145-149.

Root, T. L. and Schneider, S. H., 1993, Can Large-Scale Climatic Models Be Linked with Multi-Scale Ecological Studies? Conservation Biology, **7**:256-270.

Roughgarden, T. and Schneider, S. H., 1997, (in preparation).

Schneider, S. H., 1997a. Integrated Assessment Modeling of Global Climate Change: Transparent Rational Tool for Policy Making or Opaque Screen Hiding Value-laden Assumptions? Environmental Modeling and Assessment (submitted).

Schneider, S. H., 1997b, Laboratory Earth: The Planetary Gamble We Can't Afford to Lose. Basic Books, New York, 174 pp.

Schneider, S. H., 1996, The Future of Climate: Potential for Interaction and Surprises. In Climate Change and World Food Security . T. E. Downing, ed. NATO ASI Series: Ser. I, Global Environmental Change, vol. 37, pp. 77-113, Heidelberg: Springer-Verlag.

Schneider, S. H., 1985, Science by Consensus: The Case of the National Defense University Study 'Climate Change to the Year 2000,' An Editorial, Climatic Change, **7**(2):153-157.

Singer, S. F., 1996, Climate Change and Consensus, Science, **271**(5249):581.

Senior Seismic Hazard Analysis Committee, 1997, "Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on the Uncertainty and the Use of Experts", U. S. Nuclear Regulatory Commission, NUREG/CR-6372, Washington, D. C.

Stewart, T. R. and Glantz, M. H., 1985, Expert Judgment and Climate Forecasting: A Methodological Critique of 'Climate Change to the Year 2000,' Climatic Change, **7**(2):159-183.

Titus, James G., and Vijay K. Narayanan, 1995, Probability of Sea Level Rise, Environmental Protection Agency Report 230-R-95-008, 186 pages, October.

Titus, James G., and Vijay K. Narayanan, 1996, The Risk of Sea Level Rise, Climatic Change, **33**:151-212.

Forward to Next Section//Back to Table of Contents//AGCI Homepage//Comments: agcimail@agci.org