

# 3. IPCC's evaluation of evidence and treatment of uncertainty

The evolving nature of climate science, the long time scales involved, and the difficulties of predicting human impacts on and responses to climate change mean that many of the results presented in IPCC assessment reports have inherently uncertain components. To inform policy decisions properly, it is important for uncertainties to be characterized and communicated clearly and coherently. Since its second assessment, the IPCC has issued formal guidance for characterizing and communicating uncertainty in its reports. The guidance is intended to provide a common language for expressing confidence in the conclusions and in the likelihood that a particular event will occur.

This chapter describes how each Working Group implemented the uncertainty guidance in the fourth assessment, including how uncertainty about measurements and model results are reported and how scientific confidence in report conclusions is presented in each Summary for Policymakers and Technical Summary. This chapter also explores whether uncertainty is characterized appropriately, given the nature of IPCC assessments, and whether the scales used to characterize confidence in results are appropriate, given the nature of the conclusions. At the end of the chapter, the Committee summarizes its conclusions and recommendations for improving the presentation of evidence and treatment of uncertainty in IPCC assessment reports.

## Uncertainty guidance in the Fourth Assessment Report

IPCC authors are tasked to review and synthesize available literature rather than to conduct original research. This limits their ability to formally characterize uncertainty in the assessment reports. As a result, IPCC authors must rely on their subjective assessments of the available literature to construct a best estimate and associated confidence levels.

The IPCC guidance for characterizing uncertainty in the fourth assessment is provided in Appendix D (see 'Uncertainty guidance for the Fourth Assessment Report'). The guidance describes three approaches to indicating confidence in a particular result and/or the likelihood that a particular conclusion is correct:

1. A qualitative level-of-understanding scale (Table 3.1) describes the level of scientific understanding on a particular point in terms of the amount of evidence available and the degree of agreement among experts. There can be limited, medium, or much evidence, and agreement can be low, medium, or high. According to the guidance, when the level of confidence in the scientific findings is ‘high agreement, much evidence,’ authors may use one of the quantitative scales to calibrate the level of confidence in their conclusions or the likelihood of an outcome. The guidance also allows authors to use a quantitative scale whenever they deem it appropriate.
2. A quantitative confidence scale (Table 3.2) estimates the level of confidence for a scientific finding, and ranges from ‘very high confidence’ (9 out of 10 chance) to ‘very low confidence’ (less than 1 out of 10 chance). The Summary for Policymakers and the Technical Summary use the descriptive terminology, rather than the associated numeric value.
3. A quantitative likelihood scale (Table 3.3) is used to represent ‘a probabilistic assessment of some well-defined outcome having occurred or occurring in the future.’ The scale ranges from ‘virtually certain’ (greater than 99 percent probability) to ‘exceptionally unlikely’ (less than 1 percent probability). As in the case of Table 3.2, the Summary for Policymakers and the Technical Summary use the descriptive terminology, rather than the associated numeric value.

**Table 3.1 Level-of-understanding scale**

Level of agreement or consensus ↑	High agreement limited evidence	...	High agreement much evidence
	...	...	...
	Low agreement limited evidence	...	Low agreement much evidence
	Amount of evidence (theory, observations, models) →		

**Table 3.2 Confidence scale**

<b>Terminology</b>	<b>Degree of confidence in being correct</b>
Very high confidence	At least 9 out of 10 chance of being correct
High confidence	About 8 out of 10 chance
Medium confidence	About 5 out of 10 chance
Low confidence	About 2 out of 10 chance
Very low confidence	Less than 1 out of 10 chance

**Table 3.3 Likelihood scale used by Working Group I\***

<b>Terminology</b>	<b>Likelihood of the occurrence/outcome</b>
Virtually certain	> 99% probability of occurrence
Extremely likely	> 95% probability
Very likely	> 90% probability
Likely	> 66% probability
More likely than not	> 50% probability
About as likely as not	33 to 66% probability
Unlikely	< 33% probability
Very unlikely	< 10% probability
Extremely unlikely	< 5% probability
Exceptionally unlikely	< 1% probability

\*The 'extremely likely,' 'more likely than not,' and 'extremely unlikely' categories are not included in the IPCC guidance (Appendix D).

Each Working Group in the fourth assessment used the level-of-understanding, confidence, and likelihood scales in a different way. Working Group I relied primarily on the likelihood scale, but supplemented it with quantitative descriptions of uncertainty about outcomes—usually the endpoints of a 90 percent confidence interval or a probability distribution. Occasionally the confidence scale was used in lieu of the likelihood scale. Working Group II relied primarily on the confidence scale to indicate subjective confidence in qualitative results, and occasionally on the likelihood scale (e.g., when results were quoted from Working Group I). Working Group III relied exclusively on the level-of-understanding scale.

### **Nature of evidence and treatment of uncertainty by each Working Group**

The nature of the evidence presented, the extent to which the analysis is future-oriented, and the characterization of uncertainty varies greatly across Working Groups. For example, much of the analysis presented by Working Group I pertains to the measurement of observable quantities, such as atmospheric carbon dioxide (CO<sub>2</sub>) concentrations. In principle, it is possible to characterize the measurement and/or sampling error associated with these measurements using classical methods. A much smaller fraction of the literature assessed in the Working Group II and III reports pertains to measurement.

Models are used by all three Working Groups. Working Group I uses atmospheric and ocean general circulation models to model temperature in the recent past, with and without anthropogenic forcing, and to project future temperature, conditional on inputs from the Special Report on

Emissions Scenarios (SRES) scenarios (IPCC, 2000). General equilibrium models of the world economy are used extensively by Working Group III to project future greenhouse gas emissions, the response of emissions to policies (e.g., a carbon tax), and the costs of reducing emissions.

Uncertainty in the model parameters can be represented using sensitivity analysis or Monte Carlo analysis. If a probability distribution can be constructed over key parameters (based on data or on expert elicitation), one could sample from it to construct probability distributions of model outputs. Alternatively, key parameters can be varied one at a time (sensitivity analysis). Uncertainty regarding future model inputs (e.g., population, gross domestic product [GDP]) is often handled by running models conditional on common sets of inputs (scenarios). Indeed, the sets of assumptions about future population growth, growth in GDP and reliance on fossil fuels for the SRES scenarios were developed to facilitate the use of a common set of scenarios by Working Groups and researchers in the field.

A brief overview of the topics covered by each Working Group and the way uncertainty is characterized in the findings, particularly those presented in the Summary for Policymakers, is given below.

### *Working Group I*

The main topics covered in the Working Group I Summary for Policymakers are: (1) changes in human and natural drivers of climate, such as greenhouse gas emissions; (2) observations of climate change in the atmosphere, cryosphere, and oceans, including sea-level rise; (3) attribution of climate change; and (4) projection of climate changes over the rest of the 21<sup>st</sup> century. The first two topics deal with measurement, either direct measurement of observable quantities (e.g., surface temperature over the past 50 years) or indirect measurement (inferring historic CO<sub>2</sub> concentrations from ice cores). The last two topics use a hierarchy of models to reconstruct historic temperature or to predict temperature in the future, conditional on SRES scenarios.

Uncertainty about measured quantities is conveyed in the Summary for Policymakers by presenting a measure of central tendency and the endpoints of a 90 percent confidence interval. The measurement uncertainty is summarized based on the authors' judgment of the confidence intervals, which are based on studies reported in detail in the chapters of the Working Group report. When models are used, uncertainties are presented as the ranges of projections generated across the different

models, conditional on the SRES scenarios. Results showing uncertainty within individual models are also presented.

In addition to characterizing uncertainty using confidence intervals and probability distributions, Working Group I used a combination of the confidence and likelihood scales to characterize the certainty of their conclusions. Virtually every statement in the Summary for Policymakers is characterized using the terms employed by one of these scales. Table 3.4 illustrates the use of the likelihood scale, including the likelihood of a trend in extreme weather events in the late 20<sup>th</sup> century, the likelihood of a human contribution to that trend, and the likelihood of future trends in the 21<sup>st</sup> century, based on the SRES scenarios.

The confidence-scale terminology is also used. For example, “There is *high confidence* that the rate of observed sea-level rise increased from the 19<sup>th</sup> to the 20<sup>th</sup> century” (IPCC, 2007a, pp. 5-7). This may be contrasted with the use of the likelihood scale to make a similar statement: “. . . losses from the ice sheets of Greenland and Antarctica have *very likely* contributed to sea level rise over 1993 to 2003” (IPCC, 2007a, p. 5).

It should be emphasized that without complementary evidence such as confidence intervals and probability distributions, the use of the phrases in Table 3.4 would be an incomplete characterization of uncertainty. In other words, the quantitative scales used by Working Group I are appropriate only because they are supplemented by quantitative measures.

The quantitative scales used by Working Group I raise four additional issues:

1. It is unclear what the use of separate confidence and likelihood scales accomplishes. For example, one could have very high confidence that obtaining two sixes when rolling a pair of fair dice is extremely unlikely. But why not just say that obtaining two sixes when rolling a pair of fair dice is extremely unlikely? This suggests that the confidence scale is redundant when the likelihood scale is used, a point also made by Risbey and Kandlikar (2007).
2. It is well-documented in the literature that people interpret the terms ‘very unlikely,’ ‘likely’ etc. in Table 3.3 in different ways (Patt and Schrag, 2003; Budescu et al., 2009; Morgan et al., 2009). Specifically, the use of words alone may lead people to underestimate the probability of high-probability events and to overestimate the probability of low-probability events (see also Lichtenstein et al., 1978).

**Table 3.4 Recent trends, assessment of human influence on trends, and projections for extreme weather events for which there is a late-20<sup>th</sup> century trend (Table SMP.2, IPCC, 2007a, Cambridge University Press)**

Phenomenon <sup>a</sup> and direction of trend	Likelihood that trend occurred in late 20th century (typically post 1960)	Likelihood of a human contribution to observed trend <sup>b</sup>	Likelihood of future trends based on projections for 21st century using SRES scenarios
Warmer and fewer cold days and nights over most land areas	<i>Very likely</i> <sup>c</sup>	<i>Likely</i> <sup>d</sup>	<i>Virtually certain</i> <sup>d</sup>
Warmer and more frequent hot days and nights over most land areas	<i>Very likely</i> <sup>e</sup>	<i>Likely (nights)</i> <sup>d</sup>	<i>Virtually certain</i> <sup>d</sup>
Warm spells/heat waves. Frequency increases over most land areas	<i>Likely</i>	<i>More likely than not</i> <sup>f</sup>	<i>Very likely</i>
Heavy precipitation events. Frequency (or proportion of total rainfall from heavy falls) increases over most areas	<i>Likely</i>	<i>More likely than not</i> <sup>f</sup>	<i>Very likely</i>
Area affected by droughts increases	<i>Likely</i> in many regions since 1970s	<i>More likely than not</i>	<i>Likely</i>
Intense tropical cyclone activity increases	<i>Likely</i> in some regions since 1970	<i>More likely than not</i> <sup>f</sup>	<i>Likely</i>
Increased incidence of extreme high sea level (excludes tsunamis) <sup>g</sup>	<i>Likely</i>	<i>More likely than not</i> <sup>h</sup>	<i>Likely</i> <sup>i</sup>

a See Table 3.7 for further details regarding definitions.

b See Table TS.4, Box TS.5 and Table 9.4.

c Decreased frequency of cold days and nights (coldest 10%).

d Warming of the most extreme days and nights each year.

e Increased frequency of hot days and nights (hottest 10%).

f Magnitude of anthropogenic contributions not assessed. Attribution for these phenomena based on expert judgment rather than formal attribution studies.

g Extreme high sea level depends on average sea level and on regional weather systems. It is defined here as the highest 1% of hourly values of observed sea level at a station for a given reference period.

h Changes in observed extreme high sea level closely follow the changes in average sea level. {5.5} It is *very likely* that anthropogenic activity contributed to a rise in average sea level. {9.5}

i In all scenarios, the projected global average sea level at 2100 is higher than in the reference period. {10.6} The effect of changes in regional weather systems on sea level extremes has not been assessed.

3. The use of the likelihood scale conveys less information than a probability distribution. It should not replace ways of communicating uncertainty that convey more information when they are available. Based on a probability distribution one could say that ‘under scenario A1B it is very likely that mean global temperature will increase by at least 2° C by the end of the 21<sup>st</sup> century.’ But the distribution itself conveys this, as well as the probability of much larger mean temperature changes.

4. The likelihood scale used by Working Group I includes more categories than the likelihood scale presented in the IPCC guidance—including ‘extremely likely’ (greater than 95 percent probability), ‘more likely than not’ (greater than 50 percent probability), and ‘extremely unlikely’ (less than 5 percent probability)—introducing inconsistencies in the way likelihood is presented in the Fourth Assessment Report. Moreover, the use of overlapping categories can lead to logical inconsistencies. For example, if  $P(A) = 0.55$ , then A is ‘more likely than not’ and also ‘about as likely as not.’

### *Working Group II*

The Working Group II report begins with an examination of trends in various physical and biological measures (e.g., size of glaciers and lakes) that might be affected by climate change. Subsequent chapters deal with individual sectors—water; ecosystems; food, forests, and fiber; coastal systems; industry, settlement, and society; and human health—and eight regions of the world. Each chapter assesses current sensitivity and vulnerability to climate, future impacts and vulnerabilities, the costs of climate change, and possibilities for adaptation. The report ends with a synthesis of impacts and implications for sustainable development.

The material assessed by Working Group II includes measurements of recent and past trends in physical and biological processes that are directly linked to climate change, such as changes in the size of glaciers and timing of plant growth. In sectors such as health and agriculture, where the link to climate is mediated by other factors, the impact of past temperature on outcomes is studied using statistical models and, in the case of agriculture, also results from experiments.

A much larger portion of the Working Group II report is devoted to projecting the future impacts of climate change. Projecting climate trends and impacts is conditional on climate change and adaptation, both of which are inherently uncertain, and this uncertainty is likely to increase the farther in the future the projection is made. Different types of models are used to project future impacts. For example, projections of the impact of yield changes on world agricultural prices depend on supply and demand elasticities embedded in a model of world food markets. Consequently, model results and model uncertainty will vary, even if the same broad assumptions about future climate and adaptation are used. Best estimates from various models are often presented to show variation in the range of outcomes. Uncertainty analyses of individual models could also be presented, if available in the literature.

The extent to which results are quantified and measurement or model uncertainty is presented differs significantly across the chapters of the Working Group II report. For example, Chapter 5 (Food, Fibre, and Forest Products) presents estimates of the quantitative impacts of specific changes in temperature and precipitation on forests and agriculture, based on existing models. In contrast, Chapter 8 (Human Health) focuses on qualitative descriptions of the literature linking climate to mortality and morbidity, such as the direction of climate effects on health outcomes and the relative magnitudes of impacts.

The extent to which results are quantified also differs in the Working Group II Summary for Policymakers and the Technical Summary. The Summary for Policymakers presents quantitative information on the extent of agreement between different physical and biological trends and trends in temperature. Conclusions about observed impacts of climate on the natural and human environments and about future impacts (Sections B and C of the Summary for Policymakers) are usually stated in qualitative terms using the confidence and likelihood scales. No additional information is presented to characterize the uncertainty in the results of individual studies or to indicate the range of estimates across studies. In contrast, the Technical Summary includes more quantitative information about uncertainty. An illustrative figure in the Technical Summary (Figure TS.7 in IPCC, 2007b), for example, shows a range of estimates of the impact of temperature on yield changes by crop and latitude, with and without adaptation.

The Summary for Policymakers primarily uses the confidence scale in Table 3.2, which is intended to be used when there is 'high agreement, much evidence' in the literature. However, many of the conclusions in the 'Current Knowledge About Future Impacts' section of the Working Group II Summary for Policymakers are based on unpublished or non-peer-reviewed literature. For example, the following conclusions, each of which was based on a small number of unpublished studies, have been questioned (e.g., PBL, 2010):

Towards the end of the 21st century, projected sea-level rise will affect low-lying coastal areas with large populations. The cost of adaptation could amount to at least 5-10% of GDP. (*High confidence*; IPCC, 2007b, p. 13)

Agricultural production, including access to food, in many African countries and regions is projected to be severely compromised by climate variability and

change. The area suitable for agriculture, the length of growing seasons and yield potential, particularly along the margins of semi-arid and arid areas, are expected to decrease. This would further adversely affect food security and exacerbate malnutrition in the continent. In some countries, yields from rain-fed agriculture could be reduced by up to 50% by 2020. (*High confidence*; IPCC, 2007b, p. 13)

The use of the level-of-understanding scale (Table 3.1), rather than the confidence scale (Table 3.2), would have made clear the weak evidentiary basis for these statements.

Another issue is whether it is appropriate to use quantitative subjective probabilities when statements are qualitative in nature or imprecisely stated. Many of the 71 conclusions in the ‘Current Knowledge About Future Impacts’ section of the Working Group II Summary for Policy-makers are imprecise statements made without reference to the time period under consideration or to a climate scenario under which the conclusions would be true. Consider, for example, the statement:

In Central and Eastern Europe, summer precipitation is projected to decrease, causing higher water stress. Health risks due to heatwaves are projected to increase. Forest productivity is expected to decline and the frequency of peatland fires to increase. (*High confidence*; IPCC, 2007b, p. 14)

There is no indication about when these events are expected to occur or under what conditions. What changes in climate would give rise to these results? What is assumed about adaptation? It could be argued that, given the imprecision of the statement, it has an 80 percent chance of being true under *some* set of circumstances.

In the Committee’s view, assigning probabilities to imprecise statements is not an appropriate way to characterize uncertainty. If the confidence scale is used in this way, conclusions will likely be stated so vaguely as to make them impossible to refute, and therefore statements of ‘very high confidence’ will have little substantive value.<sup>11</sup> More importantly, the use of probabilities to characterize uncertainty is most appropriate when applied to empirical quantities (Morgan et al., 2009). The following statement may be true but should not be assigned a probability of occurrence:

---

<sup>11</sup> One could argue that the use of the phrase ‘up to’ in the statement ‘In some countries, yields from rain-fed agriculture could be reduced by up to 50% by 2020’ makes the conclusion certain to be true.

Nearly all European regions are anticipated to be negatively affected by some future impacts of climate change, and these will pose challenges to many economic sectors. (*Very high confidence*; IPCC, 2007b, p. 14)

### *Working Group III*

The main topics covered by the Working Group III Summary for Policymakers include: (1) trends in anthropogenic greenhouse emissions since 1970; (2) projected emissions to the year 2100 under various scenarios; (3) reductions in emissions in the year 2030 corresponding to various carbon prices; (4) the technical feasibility and cost of various methods of reducing greenhouse gas emissions for various sectors; and (5) estimates of the economywide costs of achieving various stabilization targets. There is also a discussion of the advantages and disadvantages of various policies for reducing greenhouse gas emissions, such as cap-and-trade systems and a harmonized carbon tax.

With the exception of historical trends in greenhouse gas emissions, all of the analyses by Working Group III rely on models of various sectors of the economy and are future-oriented. Top-down models are used to project global greenhouse gas emissions, their response to various policies, and the macro-economic costs of reaching various atmospheric CO<sub>2</sub> stabilization targets. For any geographic region, CO<sub>2</sub> emissions are the product of population, GDP per capita, energy usage per dollar of GDP, and the carbon intensity of energy. Policies such as a carbon tax are intended to reduce CO<sub>2</sub> emissions by providing incentives to lower energy per unit of GDP and to reduce the carbon intensity of energy. The response to a carbon tax depends on the costs of substituting other factors for energy and low-carbon for high-carbon fuels. It also depends upon how consumers respond to increases in costs.

All of the factors that affect CO<sub>2</sub> emissions and mitigation costs in top-down models are uncertain, and uncertainty about them increases with the length of the projection. In the long run, costs of substitution depend on advances in technology, which are highly uncertain and may themselves depend on assumptions about policies. Top-down models often use scenario analysis to make statements conditional on assumptions about variables such as the rate of population growth and the rate of growth in per capita GDP (variables that are determined primarily by factors other than climate change). However, models differ in the choice of other parameters (e.g., how responsive consumers are to energy prices) and in model structure. Variation in model results was generally handled in the

Fourth Assessment Report by presenting best estimates from different models or by showing a range of results across different model runs.

The Working Group III report also uses bottom-up models to discuss mitigation options, their costs, and policies to control them. Chapters 4-10 (IPCC, 2007c) discuss short- to medium-term options for reducing greenhouse gas emissions for seven sectors: energy supply, transportation, residential and commercial buildings, industry, agriculture, forestry, and waste management. In describing the quantity of emissions that could be reduced through different options, a distinction is made between technical potential (what can be achieved by options that are technically feasible, regardless of cost); market potential (what can be achieved by options that would pay for themselves at market discount rates and various carbon prices); and economic potential (options that would be adopted using a social rate of discount and including their social benefits—for example, reductions in local pollution from switching from coal to natural gas—as well as private benefits). The chapters also discuss policies to reduce energy consumption (e.g., energy efficiency standards for buildings, fuel taxes) and their likely effectiveness in reducing emissions.

The set of bottom-up models used to estimate emissions reductions and costs by sector is more diverse than the set of top-down models, and less detail is provided about individual model results. However, ranges of estimates are generally provided about emissions reduction potentials. In Chapter 11 (Mitigation From a Cross-Sectoral Perspective) and the Summary for Policymakers, sector-specific estimates are often aggregated to provide estimates of global mitigation potential from bottom-up studies.

The Working Group III report of the fourth assessment, including the Summary for Policymakers and the Technical Summary, relied exclusively on the use of the level-of-understanding scale (Table 3.1), indicating the amount of evidence and level of agreement about a conclusion. Three examples of the use of this qualitative scale in the Summary for Policymakers are given below:

There are multiple mitigation options in the transport sector, but their effects may be countered by growth in the sector. Mitigation options are faced with many barriers, such as consumer preferences and lack of policy frameworks (*medium agreement, medium evidence*; IPCC, 2007c, p. 13)

In 2030 macro-economic costs for multi-gas mitigation, consistent with emissions trajectories towards stabilization between 445 and 710 ppm are estimated

at between a 3% decrease of global GDP and a small increase, compared with the baseline (*high agreement, medium evidence*; IPCC, 2007c, p. 11)

Government support through financial contributions, tax credits, standard setting and market creation is important for effective technology development, innovation and deployment. Transfer of technology to developing countries depends on enabling conditions and financing (*high agreement, much evidence*; IPCC, 2007c, p. 20)

Should this scale have been supplemented with one of the quantitative scales? According to the IPCC uncertainty guidance, quantitative scales should be used when the results are themselves quantified and when there is ‘high agreement, much evidence.’ For many of the Working Group III conclusions, this is clearly not the case. Most (22 of 26) of the main conclusions in the Summary for Policymakers are qualitative, such as the first and third statements above. The use of a likelihood scale seems inappropriate for such statements. The second statement above, which is quantitative, is based on results of a suite of models that make different assumptions about the nature of technical progress, the sensitivity of consumers and producers to changes in energy prices, and other aspects of human behavior. The Summary for Policymakers and Technical Summary typically show the range of (best) estimates from a set of models when presenting a quantitative result. Different modeling assumptions provide a range of distinct point estimates of the quantity of interest, and this range should not be interpreted as being equivalent to a prediction interval for a single, correct model. Without further attempts to distinguish among models, it would be inappropriate to apply the likelihood scale to the resulting range.

Swart et al. (2009) argue that it is inappropriate for Working Group III to use the likelihood scale because of the difficulties of modeling human choice. This is also true of Working Group II’s analysis of future climate impacts and their costs, and the costs of adaptation. The uncertainty in such models is best characterized by sensitivity analyses that highlight the role of key parameters in driving model results and, when appropriate, by formal uncertainty analyses (e.g., Webster et al., 2003).

### Conclusions and recommendations

The IPCC uncertainty guidance provides a good starting point for characterizing uncertainty in the assessment reports. However, the guidance was

not consistently followed in the fourth assessment, leading to unnecessary errors. For example, authors reported high confidence in statements for which there is little evidence, such as the widely quoted statement that agricultural yields in Africa might decline by up to 50 percent by 2020. Moreover, the guidance was often applied to statements that are so vague they cannot be disputed. In these cases the impression was often left, incorrectly, that a substantive finding was being presented.

Scientific uncertainty is best communicated by indicating the nature, number, and quality of studies on a particular topic, as well as the level of agreement among studies. The level-of-understanding scale is a convenient shorthand way of communicating this information in summary documents.

### Recommendation

- ▶ Each Working Group should use the qualitative level-of-understanding scale in its Summary for Policymakers and Technical Summary, as suggested in IPCC's uncertainty guidance for the Fourth Assessment Report. This scale may be supplemented by a quantitative probability scale, if appropriate.

The IPCC uncertainty guidance urges authors to provide a traceable account of how authors determined what ratings to use to describe the level of scientific understanding (Table 3.1) and the likelihood that a particular outcome will occur (Table 3.3). However, it is unclear whose judgments are reflected in the ratings that appear in the Fourth Assessment Report or how the judgments were determined. How exactly a consensus was reached regarding subjective probability distributions needs to be documented. The uncertainty guidance for the Third Assessment Report required authors to indicate the basis for assigning a probability to an outcome or event (Moss and Schneider, 2000), and this requirement is consistent with the guidance for the Fourth Assessment Report.

### Recommendation

- ▶ Chapter Lead Authors should provide a traceable account of how they arrived at their ratings for level of scientific understanding and likelihood that an outcome will occur.

In addition, IPCC's uncertainty guidance should be modified to strengthen the way in which uncertainty is addressed in upcoming

assessment reports. In particular, quantitative probabilities (subjective or objective) should be assigned only to well-defined outcomes and only when there is adequate evidence in the literature and when authors have sufficient confidence in the results. Assigning probabilities to an outcome makes little sense unless researchers are confident in the underlying evidence (Risbey and Kandlikar, 2007), so use of the current likelihood scale should suffice.

### Recommendation

- ▶ Quantitative probabilities (as in the likelihood scale) should be used to describe the probability of well-defined outcomes only when there is sufficient evidence. Authors should indicate the basis for assigning a probability to an outcome or event (e.g., based on measurement, expert judgment, and/or model runs).

The Working Group II Summary for Policymakers in the Fourth Assessment Report contains many vague statements of ‘high confidence’ that are not supported sufficiently in the literature, not put into perspective, or are difficult to refute. The Committee believes that it is not appropriate to assign probabilities to such statements. There is, moreover, a danger that the confidence scale may be misinterpreted as indicating a statistical level of confidence in an outcome. Subjective probabilities may be assigned legitimately to well-defined outcomes using the likelihood scale. The presentation of results in the Fifth Assessment Report would be strengthened by assigning subjective probabilities *only* to well-defined conclusions.

### Recommendation

- ▶ The confidence scale should not be used to assign subjective probabilities to ill-defined outcomes.

Studies have found that individuals interpret the words in the likelihood scale very differently (Morgan et al., 2009; Patt and Schrag, 2003), even when given the terminology and probability scale in Table 3.3 (Budescu et al., 2009). An individual is more likely to understand the authors’ intent if results are stated using both probability ranges and words (Budescu et al., 2009).

## Recommendation

- ▶ The likelihood scale should be stated in terms of probabilities (numbers) in addition to words to improve understanding of uncertainty.

Studies suggest that informal elicitation measures, especially those designed to reach consensus, lead to different assessments of probabilities than formal measures. (Protocols for conducting structured expert elicitation are described in Cooke and Goossens [2000].) Informal procedures often result in probability distributions that place less weight in the tails of the distribution than formal elicitation methods, possibly understating the uncertainty associated with a given outcome (Morgan et al., 2006; Zickfeld et al., 2007).

## Recommendation

- ▶ Where practical, formal expert elicitation procedures should be used to obtain subjective probabilities for key results.

