Global Environmental Change 44 (2017) 1–14

Contents lists available at ScienceDirect

Global Environmental Change

journal homepage: www.elsevier.com/locate/gloenvcha

Unleashing expert judgment in assessment

Katharine J. Mach [a,*], Michael D. Mastrandrea [b], Patrick T. Freeman [b], Christopher B. Field [c]

[a] Department of Earth System Science, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA
[b] Department of Global Ecology, Carnegie Institution for Science, 260 Panama St., Stanford, CA 94305, USA
[c] Stanford Woods Institute for the Environment, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 17 June 2016
Received in revised form 11 February 2017
Accepted 21 February 2017
Available online

Keywords:
Expert judgment
Assessment
Uncertainties
Intergovernmental Panel on Climate Change
Fifth Assessment Report

ABSTRACT

Assessment evaluates accumulated knowledge and its limits. It informs and ideally empowers decisions and actions on complex, contested issues with persistent uncertainties. Applying rigorous expert judgment is an important dimension of assessment. Here we evaluate advances and challenges in approaches to expert judgment in the Intergovernmental Panel on Climate Change's Fifth Assessment Report (IPCC AR5). We find that revised guidance for author teams improved the development of balanced judgments on scientific evidence across disciplines. In particular, expert judgments underpinning conclusions are more extensively, transparently, and consistently communicated: degree-of-certainty terms are more abundant in AR5 policymaker summaries; wider ranges of possible outcomes are presented with greater inclusion of lower-certainty, decision-relevant findings; and expert judgments supporting conclusions are more comparable across working groups. But challenges in developing and communicating assessment conclusions persist, especially for findings with substantial uncertainties and for subjective aspects of judgments. Based on our evaluations and AR5 lessons learned, we propose a simpler, more rigorous framework for developing and communicating expert judgments in environmental assessment. We also describe practices for reducing expert-judgment biases, for advancing integration of evidence and expert judgment, and for addressing subjective dimensions of expert opinion directly and proactively.

1. Introduction: Expert Judgment in Assessment

Assessment evaluates accumulated knowledge and its limits, generally on topics transcending disciplinary boundaries and interfaces among science and societies. Through integrative assessment, experts synthesize and communicate understanding for societally important questions. The goal is to frame, support, and empower decisions on complex, contested issues with persistent uncertainties (Mitchell et al., 2006). To increase their relevance and traction, assessment processes often directly include decision-makers. Such two-way, iterative interactions enable experts to understand key decision-making questions and approaches. They can also foster decision-maker understanding of resulting assessment products, bolstering choices and actions. Prominent environmental assessments that have informed diverse public and private decisions include Intergovernmental Panel on Climate Change (IPCC) reports (e.g., IPCC, 2013a, 2014a,b,c,d), the Millennium Ecosystem Assessment (e.g., MEA, 2005), the Global Energy Assessment (GEA, 2012), the Scientific Assessments of Ozone Depletion (e.g., WMO, 2014), and many other efforts spanning international to local scales.

Rigorous expert judgment is an essential dimension of assessment. Integrating evidence across disciplines, for example, requires evaluating and synthesizing diverse research results, underpinned by different methods, assumptions, terminologies, uncertainties, and analytical strengths and weaknesses. Building from such integration, expert judgment is needed to answer policy- and decision-relevant questions, develop conclusions, and ensure effective communication of the state of scientific understanding. Participating experts must grapple with many sources of uncertainties, evaluating their implications. Sources of uncertainty can include, for instance, measurement error in observations, incomplete understanding of Earth system dynamics, and alternative approaches for modeling impacts, responses, and possible societal development trajectories. Participating experts also must consider appropriate generalizations across results, summaries

* Corresponding author. Tel.: +1 650 736 3556.
E-mail address: mach@stanford.edu (K.J. Mach).

http://dx.doi.org/10.1016/j.gloenvcha.2017.02.005

understandable for and relevant to decision-makers, and the social and cultural dimensions of assessment itself. Across such domains, effectively capturing, distilling, and conveying balanced overviews of understanding and uncertainties is not simple.

To support future environmental assessment, here we evaluate advances and challenges in approaches to expert judgment in the IPCC's Fifth Assessment Report (AR5). Over decades, the IPCC has iteratively refined methods for applying expert judgment to knowledge and uncertainties relevant to climate change (Mastrandrea and Mach, 2011). The aim has been enabling traceable, transparent expert judgments in assessment across the scientific, technical, and socioeconomic literature on climate change, its causes and impacts, and the options for response. In this tradition, the AR5 began with a revisiting and revision of the expert-judgment guidance for authors (Mastrandrea et al., 2010, 2011). In this study, we consider the degree to which the IPCC AR5 moved toward a coherent and comprehensible framework for expert judgment, evaluating successes and failures. Section 2 provides an overview of IPCC AR5's guidance for expert judgment in assessment. Sections 3 and 4 evaluate its application across working groups, also with comparison to the previous IPCC assessment report (the IPCC's Fourth Assessment Report, or AR4). Section 5 introduces possible next directions for expert judgment in environmental assessment.

2. The Revised Expert-Judgment Approach for the IPCC AR5

2.1. The IPCC AR5 Framework for Characterizing Knowledge and Uncertainties

Since its first report in 1990, IPCC assessments have included designated terms and other methods to communicate authors' expert judgments (Mastrandrea and Mach, 2011). Approaches have ranged from broad summary headings to calibrated scales for characterizing degrees of certainty in assessment conclusions. The overall goal has been to facilitate consistent treatment of uncertainties in characterizing and communicating the state of knowledge. Since the Third Assessment Report (2001), authors have worked from shared expert-judgment guidance. The 2013/2014 AR5, however, is the first IPCC report to adopt a single framework (Fig. 1) that could be applied consistently across working groups, spanning diverse disciplines and topics (Mastrandrea et al., 2010, 2011). This shared framework aimed to increase the comparability of assessment conclusions across all topics related to climate change, from the physical science basis to resulting impacts, risks, and options for response.

Building from previous guidance (Moss and Schneider, 2000; IPCC, 2005), the IPCC AR5 framework (Mastrandrea et al., 2010) conceptualizes assessment of scientific understanding and

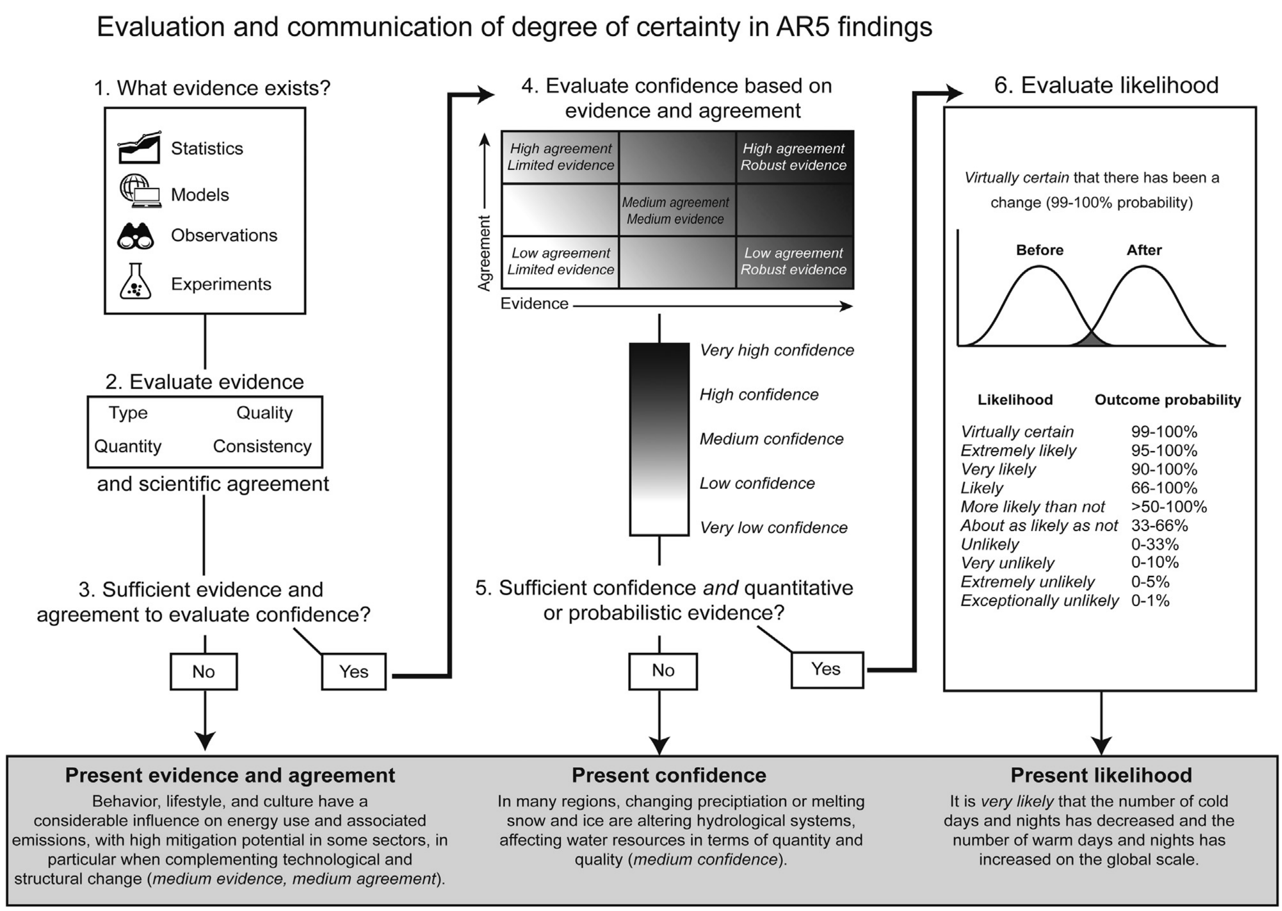

Fig. 1. The IPCC AR5 expert-judgment approach for characterizing assessment findings. This diagram illustrates the process IPCC AR5 authors used to evaluate and communicate the state of knowledge in their assessment (Mastrandrea et al., 2010). The process begins with evaluation of evidence and agreement (steps 1–3). Where possible, authors then evaluate confidence, synthesizing evidence and agreement in one qualitative metric (steps 3–5). Where uncertainties can be quantified probabilistically, authors subsequently evaluate likelihood or a more precise measure of probability (steps 5–6). Note that the likelihood categories should be considered to have "fuzzy" boundaries (step 6; CCSP, 2009; Mastrandrea et al., 2010). Unless otherwise specified, assessment conclusions characterized probabilistically are underpinned by high or very high confidence. Authors present evidence/agreement, confidence, or likelihood terms with assessment conclusions, communicating their expert judgments accordingly. Icons in step 1 were made by Freepik and Sarfraz Shoukat from www.flaticon.com; example conclusions were drawn from the IPCC AR5.

uncertainties as starting from an evaluation of **evidence** and **agreement**—in particular, the type, amount, quality, and consistency of evidence and the degree of agreement (see steps 1–3 in Fig. 1). Types of evidence can include observations, experimental results, process-based understanding, statistical analyses, or model outputs. The most robust evidence consists of multiple lines that are consistent, independent, and high quality. The degree of agreement expands upon consistency of evidence, addressing the extent of established, competing, or speculative explanations for given topics across the scientific community. The evaluation of evidence and agreement constitutes a traceable account for each assessment finding.

The next step in the IPCC AR5 framework is evaluating levels of **confidence**, which synthesize evidence and agreement in one metric (see steps 3–5 in Fig. 1). Confidence communicates qualitative judgments of the validity of findings, enabling relative comparisons across assessment conclusions. Increasing evidence and agreement is correlated with increasing confidence (step 4 in Fig. 1).

Where uncertainties can be quantified probabilistically, the IPCC AR5 framework includes a next option of characterizing assessment conclusions with **likelihood** terms or more precise presentations of probability (see steps 5–6 in Fig. 1). Likelihood can indicate probabilities for single events or broader outcomes. The associated probabilistic judgments may build from statistical or modeling analyses, elicitation of expert views, or other quantitative analyses. The framework encourages authors, where appropriate, to present probability more precisely than can be done with the likelihood scale, for example with complete probability distributions or percentile ranges, including quantification of tails of distributions important for risk management. In most cases, likelihood evaluations are underpinned by high or very high confidence in the findings.

Compared to previous IPCC guidance, the AR5 framework clarifies relations and distinctions among the three expert-judgment metrics: evidence/agreement, confidence, and likelihood (Mastrandrea and Mach, 2011). Evidence and agreement underpin confidence assignments, and confidence supports likelihood assignments. Additionally, evidence is expanded to explicitly include multiple dimensions: type, amount, quality, and consistency. The confidence and likelihood metrics are also better differentiated. Confidence is a *qualitative* metric of scientific understanding, consisting of five categories defined non-probabilistically; likelihood is a subsequent option for *quantitatively* characterizing outcome probabilities. Confidence, for example, can be applied even where deep uncertainties preclude meaningful quantification of outcome probabilities, given limited representation or ignorance of relevant processes. Overall, the goal was a harmonized, more broadly applicable approach, enabling consistent treatment of uncertainties in communicating the assessed state of knowledge.

2.2. Additional IPCC AR5 Expert-Judgment Guidance Toward Effective Assessment

Additional IPCC AR5 guidance established context for characterizing knowledge and uncertainties, extending beyond designated expert-judgment terms into practices and approaches for applying expert judgment and developing assessment findings (Mastrandrea et al., 2010, 2011).

A first realm of this guidance is appropriate generalization in assessment conclusions, emphasizing integrative, nuanced, and rigorous insights (Mastrandrea et al., 2010). A fundamental assessment challenge is communicating the overall status of knowledge while also making statements meaningfully specific, with all of the crucial detail (NEAA, 2010; IAC, 2010). For example, conclusions can be too highly generalized, either by presenting obvious truisms or by overextending beyond the scope of supporting evidence. At the other extreme, conclusions can err through overly rote relaying of specific research results, for instance without communicating judgments regarding their quality, consistency, or implications compared to other evidence lines. Guidance to authors therefore encouraged appropriate generalization across assessed literature, guarding against unsupported overgeneralization. In particular, underlying, context-rich assessment of examples across the literature needs to bolster conclusions' inferences, extrapolations, and assigned degrees of certainty. Author guidance also presented the option of "if–then" conditional conclusions that separately communicate judgments about causes and their conditional effects. For example, higher-certainty trends in climatic hazards (the "if") could be characterized separately from lower-certainty complex risks for people exposed to such hazards (the "then").

A second focal area is biases that can creep into assessment, including its group dynamics (Mastrandrea et al., 2010). Biases range from tendencies toward overconfidence and anchoring (Morgan, 2014) to inclinations toward minimizing false-positive (Type I) errors without full reflection of resultant trade-offs (Brysse et al., 2013; Anderegg et al., 2014). Guidance prioritized consideration of balanced approaches for evaluating ranges of views and developing conclusions.

Third, author guidance emphasized evaluating and communicating ranges of potential outcomes, characterizing their probabilities and consequences, to inform decision-making and risk management (Mastrandrea et al., 2010). Quantitative and qualitative information on distribution tails is important especially when low-probability outcomes may result in severe or widespread impacts. Evaluating the effectiveness of response options across ranges of climatic and socio-economic futures can also inform decisions. Emphasizing the relevant ranges of potential outcomes often requires careful judgments about evidence with large uncertainties.

Fourth, guidance to authors underscored that conclusion framings can shape interpretations (Mastrandrea et al., 2010). As one example, a reader may interpret a 10% chance of a dire outcome (e.g., a 10% chance of dying) more negatively than a reciprocal, equivalent 90% chance of a positive outcome (e.g., a 90% chance of living). As another, a reader may interpret a range of "up to 80%" (e.g., up to 80% of species affected) very differently from a more fully described range of "10–80%" (e.g., 10 to 80% of species affected). Careful, considered characterizations within conclusions can help reduce unintended interpretations and also provide more balanced information supporting risk management.

A fifth guidance point is the importance of considering all plausible sources of uncertainty (Mastrandrea et al., 2010). Such sources include difficult-to-encompass, easy-to-underestimate uncertainties arising from incomplete understanding or competing conceptual frameworks across the literature. As one implication, in some cases with quantitative evidence but deep uncertainties, confidence assignments, rather than likelihood, may better facilitate nuanced characterization of the current state of knowledge.

3. Expert Judgments Unleashed and Revealed: IPCC AR5 versus AR4

3.1. Methods of Analysis

We analyzed usage of designated terms for communicating authors' expert judgments across the IPCC AR4 and AR5. This section's analysis considers overall patterns of usage, whereas in Section 4, we more deeply evaluate expert judgments for several prominent examples. In both sections, we focus on the integrative, highly visible policymaker summaries, which undergo the IPCC's unique process of governmental review and line-by-line approval. Per IPCC rules and procedures, conclusions in a policymaker summary must trace to underlying assessment in the full report, including its chapter executive summaries and supporting sections. Although conclusions have greater visibility in the policymaker summaries, they link directly to the underlying report. For both assessment reports, the summaries for policymakers (SPMs) include the Working Group I (WGI) SPM on the physical science basis (IPCC, 2007a, 2013b); the Working Group II (WGII) SPM on impacts, adaptation, and vulnerability (IPCC, 2007b, 2014e); the Working Group III (WGIII) SPM on mitigation (IPCC, 2007c, 2014f); and the Synthesis Report (SYR) SPM (IPCC, 2007d, 2014g).

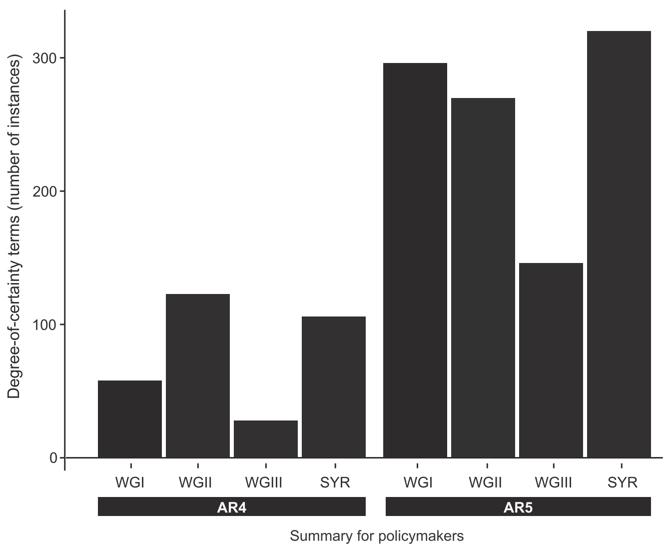

Fig. 2. Usage of degree-of-certainty terms in the eight policymaker summaries of the IPCC AR4 and AR5. Terms include instances of evidence/agreement, confidence, and likelihood (IPCC, 2005; Mastrandrea et al., 2010).

For each SPM, we tabulated unique instances of evidence/agreement, confidence, and likelihood terms (IPCC, 2005; Mastrandrea et al., 2010). We included terms presented within SPM text, boxes, footnotes, figures, tables, and captions. We excluded introductory usage not characterizing assessment findings (e.g., where the IPCC expert-judgment framework is introduced and terms defined). A unique instance of evidence/agreement terms encompassed both the level of evidence and the degree of agreement (i.e., "*robust evidence*, *high agreement*" is one instance) except where an SPM included unmatched evidence or agreement terms. Uses per page were based on printed layout, excluding title pages.

We also evaluated certainty levels across SPM evidence/agreement and confidence terms. Given sparse usage of the "*very low*" and "*very high*" confidence levels, we grouped confidence levels into three categories: low certainty (*very low* and *low confidence*), medium certainty (*medium confidence*), and high certainty (*high* and *very high confidence*). Evidence/agreement combinations were classified as follows: "*limited evidence*, *low agreement*," "*limited evidence*, *medium agreement*," and "*medium evidence*, *low agreement*"—low certainty level; "*medium evidence*, *medium agreement*," "*robust/much evidence*, *low agreement*," and "*limited evidence*, *high agreement*"—medium certainty level; "*robust/much evidence*, *high agreement*," "*robust/much evidence*, *medium agreement*," and "*medium evidence*, *high agreement*"—high certainty level. Note that the AR4 guidance (IPCC, 2005) uses the term "*much evidence*" instead of the AR5 framework's "*robust evidence*." For this certainty-level analysis, we excluded likelihood terms because they characterize outcome probabilities rather than levels of scientific understanding. We also excluded unmatched evidence or agreement terms.

To compare SPM degree-of-certainty terms to the underlying assessment, we additionally tabulated unique instances of evidence/agreement, confidence, and likelihood terms across the 30 chapter executive summaries in the WGII AR5 (IPCC, 2014a, b).

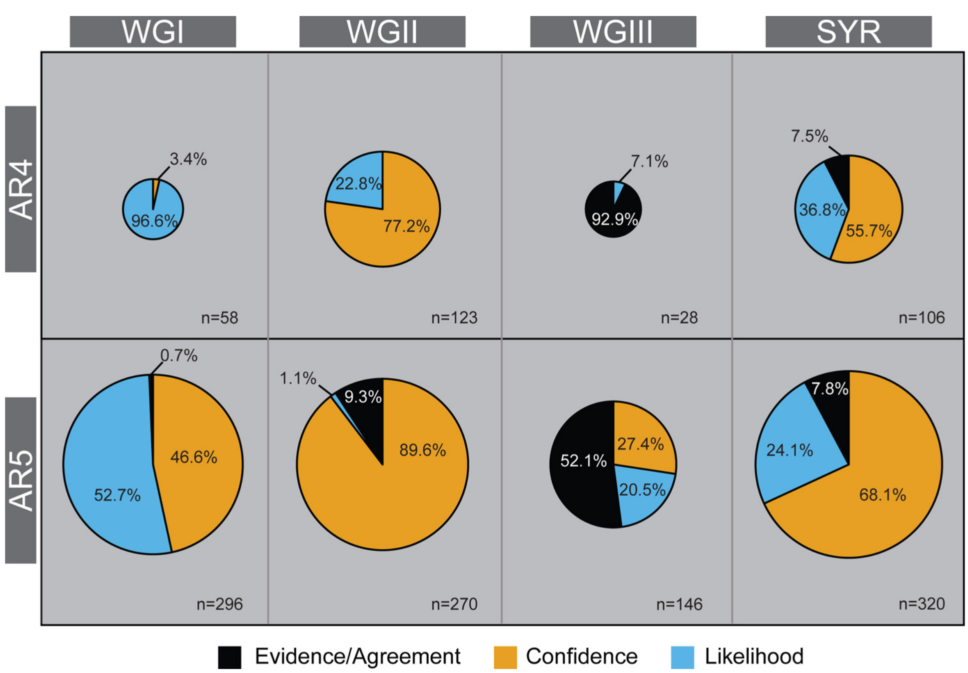

Fig. 3. Proportional usage of evidence/agreement, confidence, and likelihood terms across the eight policymaker summaries of the IPCC AR4 and AR5. For each SPM, chart area is proportional to the SPM's overall number of degree-of-certainty terms, as noted in the bottom right of each panel. Usage of each scale is depicted with a different color, indicating percentage of overall SPM degree-of-certainty terminology.

Counts and certainty levels were evaluated for executive summaries as for the SPMs.

Occurrence of degree-of-certainty terms, including numbers of instances and certainty levels, was compared across AR5 versus AR4 SPMs with the Mann-Whitney U-test, which is a non-parametric alternative to a two-sample t-test, and with Fisher's Exact test, which outperforms the chi-squared test when expected numbers are small (McDonald, 2014).

3.2. Results and Discussion of IPCC AR5 and AR4 Expert Judgments

3.2.1. Expert Judgments across AR5 versus AR4 Policymaker Summaries

The IPCC AR5 SPMs, compared to the AR4 SPMs, use substantially more of the designated terms for communicating authors' expert judgments (Fig. 2) (Mann-Whitney U-test, $p = 0.03$). The total number of degree-of-certainty terms increases from 315 instances across AR4 SPMs to 1032 across AR5 SPMs. Terms increase even accounting for the greater length of AR5 SPMs (Mann-Whitney U-test, $p = 0.06$). All AR5 SPMs were longer than the AR4 SPMs. Taking SPM length into account, normalized degree-of-certainty usage is 3.41 (WGI), 8.20 (WGII), 1.33 (WGIII), and 5.05 (SYR) terms per page in the AR4 and 11.38 (WGI), 9.00 (WGII), 5.41 (WGIII), and 10.67 (SYR) terms per page in the AR5. WGII usage changes the least on this per-page basis, given an especially short AR4 SPM and an especially long AR5 SPM. The AR5's notably increased usage of degree-of-certainty terms suggests thorough integration of the revised expert-judgment approach. This overall pattern, however, cannot be used to evaluate the quality of associated judgments, for example as considered in Section 4.

Relative usage of evidence/agreement, confidence, and likelihood varies across AR4 and AR5 SPMs (Fig. 3). For each SPM, chart area in Fig. 3 is proportional to the SPM's overall number of degree-of-certainty terms (as in Fig. 2), with colors indicating the three separate scales. Each AR4 working group (WGI, WGII, and WGIII) favors a different scale, and no working group uses all three. Note, however, that the SYR AR4 SPM does use all three scales, since it integrates findings across working groups. This AR4 pattern and its underlying causes hindered comparisons across conclusions, both within and across working groups and their different climate-change topics. The resulting challenges for both readers and participating experts motivated revision of the IPCC's expert-judgment guidance for the AR5 (IAC, 2010; Mastrandrea and Mach, 2011).

In contrast, the AR5 SPMs each use, at least to some extent, all three metrics: evidence/agreement, confidence, and likelihood (Fig. 3). Across AR5 SPMs, reliance on confidence terms increases compared to the AR4, and usage of likelihood mostly decreases. This increased reliance on confidence likely resulted from the AR5 guidance's substantial clarification of the relationships among the three scales (Section 2.1). That is, confidence could now more coherently be used together with likelihood, as particularly relevant for the WGI AR5, and with evidence/agreement, as particularly relevant for the WGIII AR5.

The working groups still favor different expert-judgment scales, but the differences more directly reflect different evidence bases across the working groups. For example, likelihood requires that uncertainties can be quantified probabilistically, with reasonable understanding of factors encompassed versus not encompassed in supporting analyses. The WGI AR5 SPM abundantly uses likelihood terms given the prevalent quantitative evidence and more quantifiable uncertainties in its scope (e.g., for climate observations, statistical detection and attribution analyses, or climate model projections). But in contrast to the AR4, the WGI AR5 SPM also uses many confidence terms, which are now more clearly linked to and distinguished from likelihood (Section 2.1), to characterize the validity of assessment findings.

The WGII AR5 SPM relies predominantly on confidence terms to communicate expert judgments (Fig. 3). This emphasis, which was

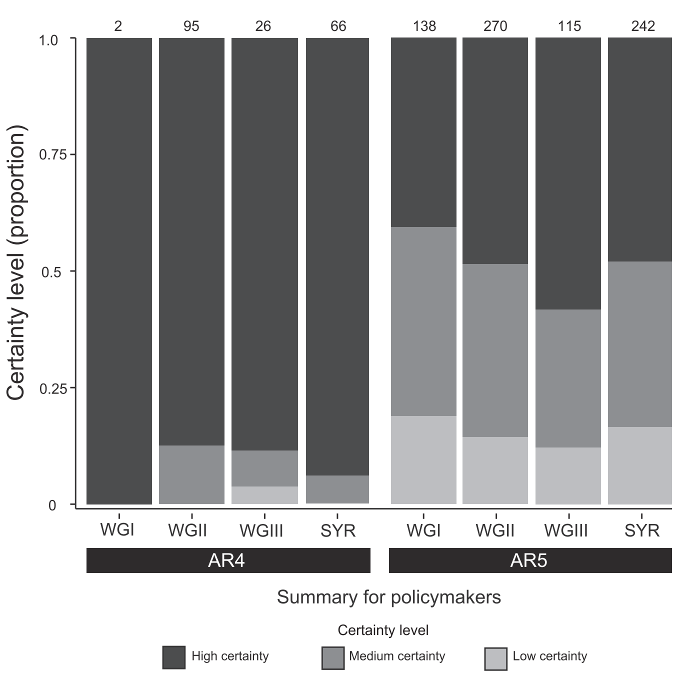

Fig. 4. Certainty levels across evidence/agreement and confidence terms in the eight policymaker summaries of the IPCC AR4 and AR5. Evidence/agreement and confidence terms are binned into low, medium, and high certainty categories as described in Section 3.1. The total number of terms across certainty levels is shown at the top of the stacked bars for each SPM.

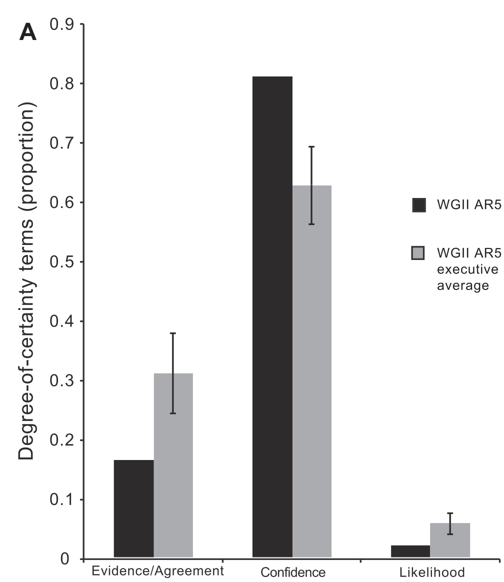

Fig. 5. Proportional usage of evidence/agreement, confidence, and likelihood terms in the WGII AR5. (A) Proportional usage of degree-of-certainty scales in the WGII AR5 SPM compared to the underlying chapter executive summaries. For the WGII AR5 chapter executive summaries, proportional usage is shown as mean ± standard error. Terms reported for the SPM exclude terms in a supplementary table that does not draw from chapter executive summaries. (B) Proportional usage of degree-of-certainty scales in WGII AR5 chapter executive summaries, depicted chapter by chapter. The total number of terms across metrics is shown at the top of the stacked bars for each chapter executive summary. Chapters 1 and 2 are introductory framing chapters; chapters 3–13 are sectoral chapters; chapters 14–17 focus on adaptation; chapters 18–20 are integrating synthesis chapters; and chapters 21–30 are regional chapters. Chapters 1–20 constitute *Part A: Global and Sectoral Aspects* of the WGII AR5, and chapters 21–30 constitute *Part B: Regional Aspects*.

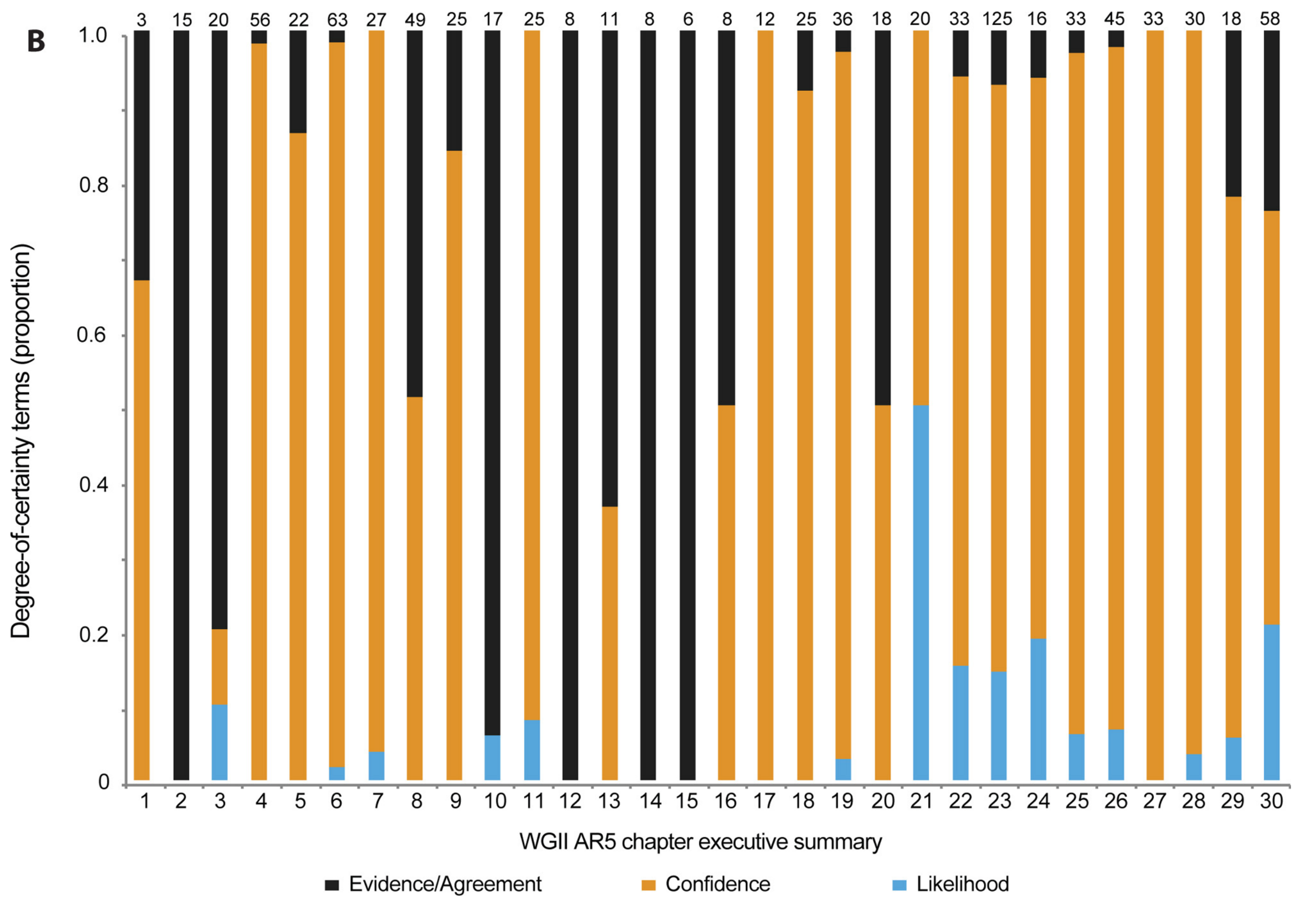

Fig. 5. (*Continued*).

explicitly encouraged in working-group-specific guidance to authors, emerged from the wide sweep of topics and disciplines under WGII purview (Mastrandrea et al., 2011). Topics span freshwater resources, ecosystems, food systems, and human security, for example, and disciplines range from glaciology to economics. The full evidence base for any topic often includes quantitative and qualitative evidence that must be holistically assessed, with uncertainties difficult to quantify across the whole scope of relevant literature. From the AR4 to the AR5, WGII therefore maintained and even increased its strong focus on confidence in communicating expert judgments. The increased incidence of evidence/agreement likely resulted from the AR5 guidance's more ambitious embedding of that scale within the overall expert-judgment approach (Section 2.1). Additionally, the reduced usage of likelihood may have followed from increased emphasis on applying the scale only where uncertainties could be meaningfully quantified and on generalizing findings at appropriate levels, for example not implying false precision where available quantitative results exclude processes known to be important (Section 2.2).

The WGIII AR5 SPM uses predominantly evidence/agreement, but with notable occurrence of confidence and likelihood terms. Likelihood is used, for example, to characterize results from integrated assessment models. The evidence/agreement emphasis may reflect some path dependence in approach, following from the WGIII AR4 preference for the scale (Mastrandrea and Mach, 2011). For instance, AR5 participating experts may have been able to assign confidence levels in some cases but instead preferred evidence/agreement, given past practices and established norms shared across individuals and disciplines, report to report.

Overall, usage of evidence/agreement, confidence, and likelihood in AR5 versus AR4 SPMs demonstrates progress toward a more harmonized and consistent approach across working groups (Figs. 2 and 3). The AR5 guidance clarified the relationships among the different metrics (Section 2.1), and as a result, each working

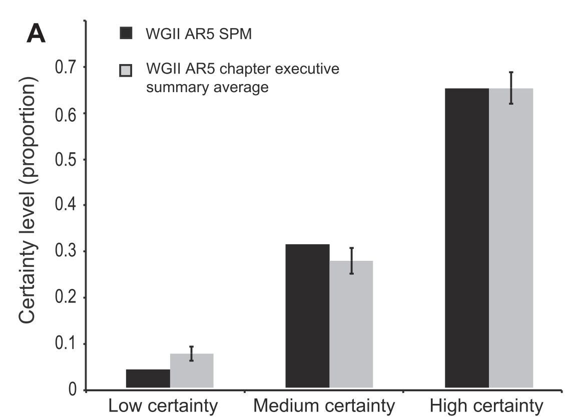

Fig. 6. Certainty levels across evidence/agreement and confidence terms in the WGII AR5. Evidence/agreement and confidence terms are binned into low, medium, and high certainty categories as described in Section 3.1. (A) Proportional occurrence of low, medium, and high certainty levels in the WGII AR5 SPM compared to the underlying chapter executive summaries. For the WGII AR5 chapter executive summaries, proportional occurrence is shown as mean ± standard error. Terms reported for the SPM exclude terms in a supplementary table that does not draw from chapter executive summaries. (B) Proportional occurrence of low, medium, and high certainty levels in WGII AR5 chapter executive summaries, depicted chapter by chapter. The total number of terms across certainty levels is shown at the top of the stacked bars for each chapter executive summary.

group could apply all three scales, matching their emphases to evidence lines and uncertainty sources in their scope. Different patterns of usage persist, but with greater comparability, intentionality, and rigor of expert-judgment approaches across SPMs and their assessment conclusions.

The certainty levels of SPM conclusions also differ across AR4 versus AR5 SPMs (Fig. 4). Low and medium certainty findings are substantially more numerous in the AR5 (Fisher's Exact tests, $p < 0.001$). Compared to the AR4, AR5 expert-judgment guidance to authors notably increased emphasis on communicating wide ranges of possible outcomes, not just high certainty conclusions (Section 2.2). The goal was developing assessment findings that could better inform risk management and decision-making under uncertainty, building from the full and growing evidence basis. Observed trends (Fig. 4) suggest that the author guidance did result in communication of wider ranges of possible outcomes, expanding beyond earlier focus on high certainty conclusions.

3.2.2. Expert Judgments in a Policymaker Summary Compared to Its Supporting Assessment

For the working-group SPMs, most SPM findings trace directly to executive-summary conclusions in the underlying chapters. We therefore analyzed, for the WGII AR5, expert-judgment trends in supporting chapters versus the integrating SPM. In the WGII AR5 SPM, proportional usage of confidence is greater than in the underlying chapter executive summaries, and proportional usage of evidence/agreement is lower (Fig. 5A). This trend reflects an intentional SPM emphasis on communicating expert judgments with confidence levels where possible, especially where conclusions trace to multiple underlying conclusions using both evidence/agreement and confidence. The confidence emphasis aimed to streamline expert-judgment characterizations for readers, increasing their accessibility with the simpler five-point scale.

In contrast, WGII AR5 SPM proportional usage of likelihood is lower than in the underlying chapter executive summaries (Fig. 5A). This trend occurs because most likelihood terms in underlying executive summaries appear in regional chapters (chapters 21–30) reporting or extending WGI conclusions about physical climatic changes (Fig. 5B). These physical-science-basis conclusions are essential to the WGII regional chapters but largely out of scope for the WGII SPM.

Most WGII AR5 chapters rely predominantly on the confidence metric for characterizing expert judgments (Fig. 5B). WGII AR5 chapters favoring evidence/agreement, including five chapters using it exclusively, draw heavily from social-science disciplines in their supporting literature bases (Fig. 5B). But chapter patterns also reveal preferences of individual author teams, for example with two chapters drawing from economics (chapters 10 and 17) favoring different scales (chapter 10: evidence/agreement; chapter 17: confidence). Overall, these WGII patterns demonstrate progress toward a harmonized expert-judgment approach encompassing needs across disciplines, topics, and uncertainty sources (Section 2.1), but still with some choices more arbitrarily decided based on preferences of individual author groups.

Compared to the WGII AR4 SPM, the WGII AR5 SPM uses substantially more evidence/agreement and confidence terms with low and medium certainty levels (Fig. 4). As compared to the underlying chapters, the WGII AR5 SPM uses a slightly lower proportion of low-certainty terms and a slightly higher proportion of medium-certainty terms (Fig. 6A). (Note that Fig. 4 includes a supplementary WGII AR5 SPM table excluded from Fig. 6A, as described in the Fig. 6A caption.) These trends suggest that the

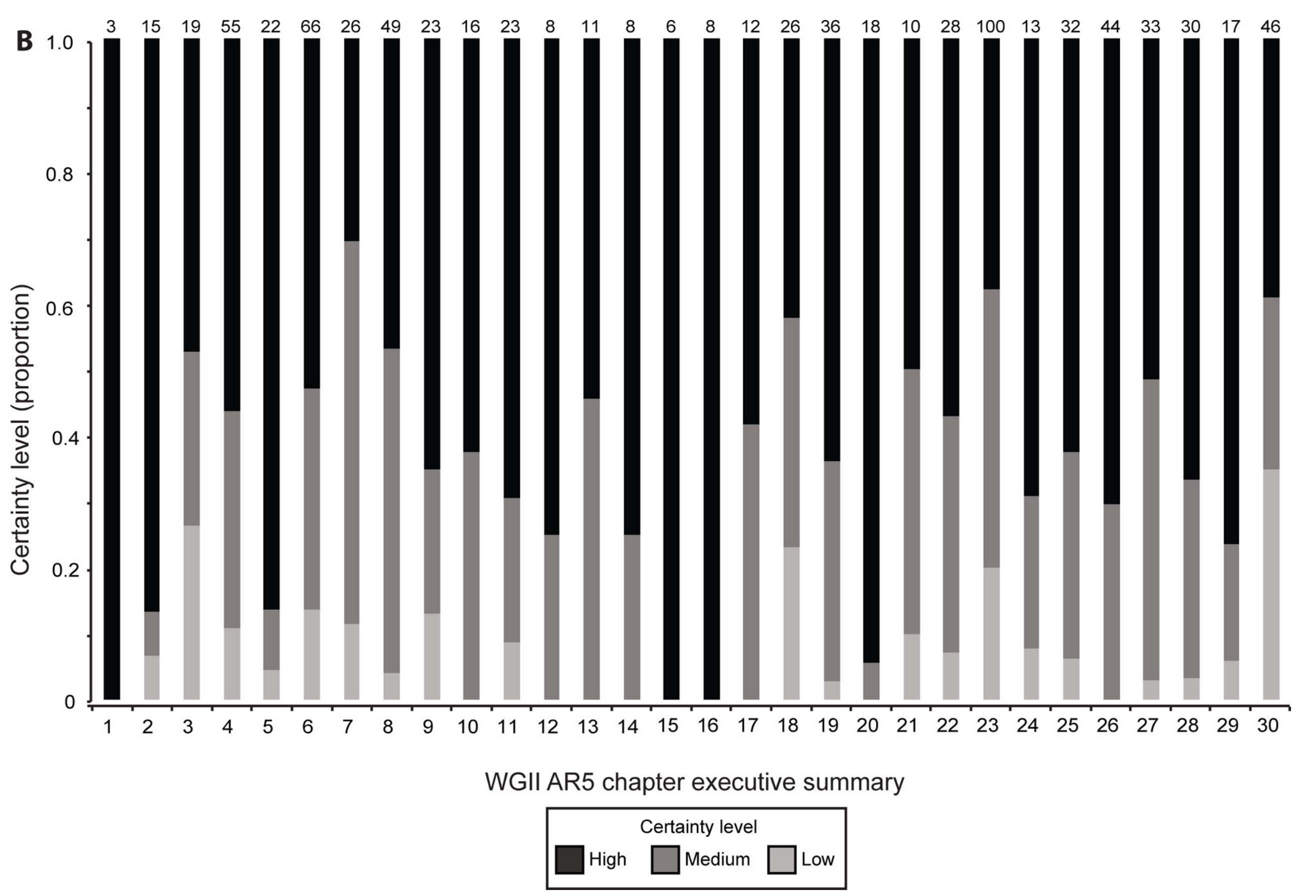

Fig. 6. (*Continued*).

increased AR5 emphasis on communicating wide ranges of possible outcomes (Section 2.2) may have been even more prominent in WGII AR5 chapters' supporting assessment as compared to the WGII AR5 SPM. Across chapter executive summaries, three chapters, including an introductory chapter, communicate solely high-certainty conclusions (Fig. 6B). All other chapters present some, and often many, conclusions characterized by low and medium certainty levels.

4. Important but Challenging IPCC AR5 Expert Judgments

Building from Section 3's overarching analysis, here we consider specific examples of important, yet challenging topics assessed in the IPCC AR5, evaluating how degree-of-certainty terms were used to communicate the current state of knowledge. We evaluate choices made in characterizing understanding and uncertainties through the AR5 approach (Section 2.1), in light of the broader themes emphasized for effective assessment (Section 2.2). In particular, we analyze two areas that entail particularly difficult choices: (1) approaches to communicating expert judgments for lower-certainty findings, and (2) interactions between subjective judgments inherent in confidence and likelihood assignments.

As compared to AR4 SPMs, AR5 SPMs include substantially more medium and low certainty findings (Section 3.2.1). These findings cover wider ranges of possible outcomes relevant to risk management. They also represent challenging territory for expert judgment, for which future assessments have compelling opportunities to make advances.

Some lower certainty findings result directly from limitations in available evidence. For instance, the WGI AR5 SPM specifies that "*Confidence* in precipitation change averaged over global land areas since 1901 is *low* prior to 1951 and *medium* afterwards" (IPCC, 2013b). In this finding, the low and medium confidence levels, linked to uncertainties resulting from limited observational data, function as a substitute for a substantive conclusion: the reader can't tell if existing lower-confidence evidence indicates precipitation increases, decreases, both, or neither (or where data are absent), aside from a later mention of mid-latitude Northern Hemisphere land-area trends. In this way, the conclusion fails to present meaningful information on the range of possible trends to date (Section 2.2). The assessment authors potentially had opportunities for communicating such decision-relevant, lower-confidence information but instead opted to relay little in the SPM conclusion, perhaps retreating to a more culturally comfortable minimization of asserted expert judgments that could prove incorrect (Brysse et al., 2013). Governmental interest in this information led to addition of a two-panel figure depicting observed precipitation change from 1901 to 2010 and from 1951 to 2010 during the SPM governmental approval session (see Figure SPM.2 in IPCC (2013b)). The figure addition perhaps underscores decision-making relevance of the full range of understanding, even where uncertainties preclude definitive assertions (Section 2.2).

Other lower-certainty findings emerge more broadly from the complexity of interactions shaping an outcome. Large numbers of interactions can result in uncertainties about which driving factors will be dominant, when, and how; this is especially the case when key factors have inherently unpredictable components. Resulting lower-certainty conclusions can pertain to physical or ecological systems. For example, the WGI AR5 SPM states, "Current estimates indicate that the threshold [for near-complete loss of the Greenland ice sheet over a millennium or more] is greater than about 1 °C (*low confidence*) but less than about 4 °C (*medium confidence*) global mean warming with respect to pre-industrial" (IPCC, 2013b). As another example, the WGII AR5 SPM indicates that "Examples [of abrupt and irreversible regional-scale ecosystems change] that could lead to substantial impact on climate are the boreal-tundra Arctic system (*medium confidence*) and the Amazon forest (*low confidence*)" (IPCC, 2014e). For these potential physical and ecological tipping points, lower-certainty confidence levels are put to better use, providing substantive conclusions for high-risk outcomes (Section 2.2). Still, future assessments could go further in communicating expert judgments saliently in such domains, for example through eliciting relative probabilities or characterizing key early warning signs to monitor.

Other lower-certainty conclusions characterize complex interactions in human systems. For example for violent conflict, the WGII AR5 SPM contends with divergent quantitative and qualitative evidence by adopting a qualitative conclusion with broad indication of mechanisms: "Climate change can indirectly increase risks of violent conflicts in the form of civil war and inter-group violence by amplifying well-documented drivers of these conflicts such as poverty and economic shocks (*medium confidence*)" (IPCC, 2014e). The assessment authors, for instance, had to weigh climate–conflict quantifications not encompassing all potentially relevant influences such as future adaptation potential, along with qualitative case studies lacking quantification of uncertainties for future risks. The conclusion adopted reflects carefully considered choices about appropriate generalization across evidence lines (Section 2.2). For economic impacts, the WGII AR5 SPM presents quantitative results with heavy qualification: "With…recognized limitations, the incomplete estimates of global annual economic losses for additional temperature increases of ~2 °C are between 0.2 and 2.0% of income (±1 standard deviation around the mean) (*medium evidence*, *medium agreement*). Losses are *more likely than not* to be greater, rather than smaller, than this range (*limited evidence*, *high agreement*)" (IPCC, 2014e). For the economic-impacts example, the authors indicate a greater than ~15% probability of losses greater than 2% of income for additional warming of ~2 °C. But in this conclusion and a subsequent statement for additional 3 °C warming, little information is provided on the potential magnitudes of exceedance, the reliability of available estimates, or the implications of omitted processes. Even with its qualifications, the SPM finding's implied precision may contrast more nuanced, comprehensive characterizations of current knowledge, which might require more qualitative generalizations (Section 2.2). Moving forward, complex risks for social and economic systems, as well as physical and ecological systems, will demand sophisticated responses by societies, increasing the imperative for assessment to grapple with diverse mechanisms and factors influencing outcomes and their likelihoods, even where uncertainties are deep.

In the AR5, quantitative evidence, especially model ranges, is often subjectively interpreted and adjusted to produce probabilistic likelihood assignments, with the adjustment accounting for unquantified uncertainties. The degree of adjustment varies across conclusions, interacts with confidence levels, and is only rarely explicit and transparent. This dynamic challenges interpretation of the AR5's expert-judgment approach (Section 2.1) in some cases. As one example, for WGI AR5 temperature and sea-level-rise projections, 5–95% model ranges are used as likely ranges. 5–95% model ranges could conceivably represent very-likely ranges if the models were judged to include all relevant uncertainties; instead, using 5–95% ranges as likely ranges implies a subjective adjustment based on confidence in the validity of the underlying models. The resulting conclusions about temperature projections include, "Increase of global mean surface temperatures for 2081–2100 relative to 1986–2005 is projected to *likely* be in the ranges derived from the concentration-driven CMIP5 model simulations, that is, 0.3 °C to 1.7 °C (RCP2.6),…2.6 °C to 4.8 °C (RCP8.5)" (IPCC, 2013b). High confidence underpins this conclusion. In contrast, for sea level rise, 5–95% model ranges are reported as likely ranges with medium confidence: "Global mean sea level rise for 2081–2100 relative to 1986–2005 will *likely* be in the ranges of 0.26 to 0.55 m for RCP2.6...and 0.45 to 0.82 m for RCP8.5 (*medium confidence*)" (IPCC, 2013b). Confidence in sea-level-rise conclusions may be lower given discrepancies across evidence lines, such as larger semi-empirical model projections assigned low confidence. Divergent evidence and substantial associated uncertainties have challenged sea-level-rise assessment, but with advances from the AR4 to the AR5 and with notable developments in the literature even since this last assessment. Across temperature and sea-level-rise projections, the implicit adjustment of using 5–95% model ranges as likely ranges is different: the sea-level-rise adjustment indicates medium confidence for example while excluding consideration of semi-empirical results in assessed probability ranges, which adds another incompletely transparent layer of judgment. Such adjustment is consistent with the AR5 guidance to authors, although its characterization leaves open questions about the range of associated individual and group expert judgments. In Section 5, we unfold potential future approaches for more transparently addressing and revealing such subjective dimensions of assessment.

Perhaps most challenging for expert judgments about future-outcome likelihoods are the deeper uncertainties around human choice. In some instances, the response has been to limit expert-judgment characterizations to part of a conclusion, sometimes the less important part. For example, in analyzing costs of mitigation pathways, WGIII AR5 SPM authors provide a likelihood for warming but not costs: "Under. . .assumptions [of immediate mitigation by all countries, a single global carbon price, and availability of all key technologies] mitigation scenarios that are *likely* to limit warming to below 2 °C through the 21st century relative to pre-industrial levels entail losses in global consumption—not including benefits of reduced climate change as well as co-benefits and adverse side effects of mitigation—of 1 to 4% (median: 1.7%) in 2030, 2 to 6% (median: 3.4%) in 2050 and 3 to 11% (median: 4.8%) in 2100 relative to consumption in baseline scenarios that grows anywhere from 300% to more than 900% over the century" (IPCC, 2014f). The structure of SPM statements presented implies high confidence for this conclusion (IPCC, 2014f, g). Uncertainties about the veracity of economic, technological, policy, and other assumptions in the scenarios, however, greatly expand the range of potential costs across possible mitigation futures, contrasting the precision of numbers reported (IPCC, 2014f). Moving forward, there will be opportunities to add value through meaningful generalizations across such quantitative results (Section 2.2), for example in assessing assumptions underpinning phenomenally ambitious 1.5 °C scenarios and attendant consequences for the global response to the threat of climate change. As the AR5's implied expert-judgment approach suggests, the challenge of encompassing all relevant uncertainties in confidence and likelihood assignments may be largest where hard-to-predict human choices, developments, and advances shape outcomes. Real-time relevance of resulting assessment, however, heightens the importance of innovating approaches for supporting decisions, despite the deep uncertainties (e.g., Drouet et al., 2015).

Across these lower-certainty and subjective-adjustment examples, assessment authors grappled with diverse uncertainties, adopting different approaches to applying expert judgment in synthesizing evidence and developing conclusions. Moving forward, assessment efforts have opportunities to apply more deliberate and transparent expert-judgment approaches for such challenging, important topics, learning from past experiences. Recognizing and communicating the implications of different expert-judgment approaches can increase the rigor, nuance, and utility of assessment conclusions that result.

5. Next Directions for Expert Judgment in Assessment

5.1. Toward a Simple, Rigorous Expert-Judgment Framework for Assessment Conclusions

Expert judgment will continue to be fundamental in environmental assessments across scales, contexts, and topics. To inform future efforts, we evaluated the IPCC AR5 framework (Fig. 1) for organizing and communicating expert judgments underlying assessment findings. The revision of the AR5 framework partially addressed the AR4's divergent patterns of degree-of-certainty term usage (Fig. 3). The AR5 framework established confidence as a qualitative metric synthesizing evidence and agreement, with likelihood as a subsequent quantitative scale for characterizing outcome probabilities (Section 2.1). Compared to previous IPCC approaches, the AR5 framework succeeded in enabling substantially more characterization of knowledge and uncertainties through designated terminology (Fig. 2). The framework was also used more consistently across IPCC working groups spanning diverse topics and disciplines (Fig. 3).

Even with these adjustments, the IPCC AR5 expert-judgment framework leaves room for improvement. The AR5 framework constructively responded to path dependencies of past IPCC approaches, to distinct working-group preferences and norms, and to diverse disciplinary differences across climate-change domains. But the framework that resulted met these challenges by cementing a complex approach (Section 2.1). This complexity enabled past IPCC expert-judgment scales to coexist more coherently in a new framework. Although an advance, the complex "compromise" framework is also challenging for readers to understand and for assessment authors to apply consistently.

Two aspects, in particular, could be revisited in advancing a simple, rigorous expert-judgment framework for future applications. First, the evaluation of confidence (Fig. 1, step 4) only incrementally extends beyond the evaluation of evidence and agreement. The five integrative confidence terms (*very low* to *very high*) are simpler and more intuitive than the nine evidence/agreement combinations (see also Section 3.2.2). But this gain is in many ways outweighed by the added complexity for a reader, given reliance on both evidence/agreement and confidence terms across conclusions. Preferences for evidence/agreement versus confidence are often shaped by somewhat arbitrary working-group and disciplinary preferences (Section 3), rather than clear added value from having two sets of terms rather than one.

Second, the relationship between confidence and likelihood, although clarified as compared to the IPCC AR4 (Mastrandrea and Mach, 2011), remains somewhat ambiguous (Aven and Renn, 2015). In practice and perhaps out of necessity, author teams adopted a spectrum of approaches in interpreting the relationship between confidence and likelihood. Redundancies and interactions across layers of characterized uncertainties were often handled differently and not transparently (Section 4). That is, in the AR5 framework, confidence is a qualitative metric of scientific understanding underpinning probabilistic likelihood assignments; the metrics characterize different aspects of uncertainties. Despite this clarification, ambiguity remains when authors adjust assessed likelihoods subjectively as a function of confidence to different degrees (Section 4). In some cases, likelihood indicates the subjective probabilistic assessment of an author team, adjusted based on confidence in the underlying evidence (e.g., an assessed probability reflecting value uncertainties and deeper uncertainties)

(Mastrandrea et al., 2011). In other cases, likelihood more directly indicates statistical or probabilistic output of models or analyses (e.g., value or parameter uncertainties), with the author team separately characterizing its confidence in the validity or correctness of the models or analyses (e.g., model, structural, or deep uncertainties).

Other recent environmental assessments have applied related, but substantially simplified, frameworks for capturing and communicating expert judgments. For example, the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services's pollination report (IPBES, 2016) adopts four evidence/agreement summary terms to characterize confidence in assessment findings: *inconclusive*, *unresolved*, *established but incomplete*, and *well established*. As in the IPCC framework (Fig. 1, step 4), evidence/agreement terms are presented in a matrix, with increasing certainty shaded along the diagonal. Brief descriptions of underlying evidence/agreement are also provided. They range from *well established* requiring comprehensive syntheses or multiple independent evidence lines to *inconclusive* representing limited evidence and major knowledge gaps. As other examples, the second US National Climate Assessment (Karl et al., 2009) used a likelihood scale to express expert judgments about the probabilities of different outcomes, while the third US National Climate Assessment (Melillo et al., 2014) primarily used confidence levels within traceable accounts for key messages but not within the key messages themselves. For assessment findings across these examples, expert judgments can be presented on a single scale. Simplifying communication of expert judgments could advance the accessibility of the IPCC's framework as well, while simultaneously increasing its rigor by reducing ambiguities described above.

Based on our analysis and experience with the IPCC AR5 expert-judgment framework, we propose an iterative improvement for future assessments (Fig. 7). This framework is intended to be simultaneously more rigorous and accessible—more straightforward for experts to apply and for readers to understand. The framework includes a single scale for describing scientific understanding (Fig. 7, step 2) based on evidence and agreement. It also includes a likelihood scale for subsequent characterization of subjective probabilistic judgments, reflecting all plausible uncertainty sources. The qualitative scientific-understanding scale, as in the IPCC AR5, can apply to evidence across disciplines. The five summary terms, for example, can characterize lower-certainty conclusions or broad qualitative conclusions if the available evidence does not support subjective probabilities. Further, we recommend more explicit description of likelihood assignments as subjective probabilistic assessments integrating collective expert judgment and available evidence. In this approach, likelihood involves evaluating all relevant uncertainties (e.g., value, structural, *and* deeper uncertainties) in defining the assessed probability of an outcome, the value of an uncertain quantity, or the state of the present or future world (Moss and Schneider, 2000; Morgan, 2014). Here, likelihood is informed by all available evidence, whether it is quantitative, probabilistic, or more diverse. Where possible and appropriate, experts would assign likelihood or more precise presentations of probability. Scientific-understanding terms would be a supplement or, when probability cannot be evaluated, a fallback. Likelihood could be prioritized especially for key assessment findings, perhaps with more abundant use of scientific-understanding terms in underlying traceable and transparent accounts of evidence and expert judgments. Accompanying practices, described in Section 5.2, are important for systematically and consistently documenting the range of expert judgments and subjective probabilities, as well as their integration, in implementing this framework.

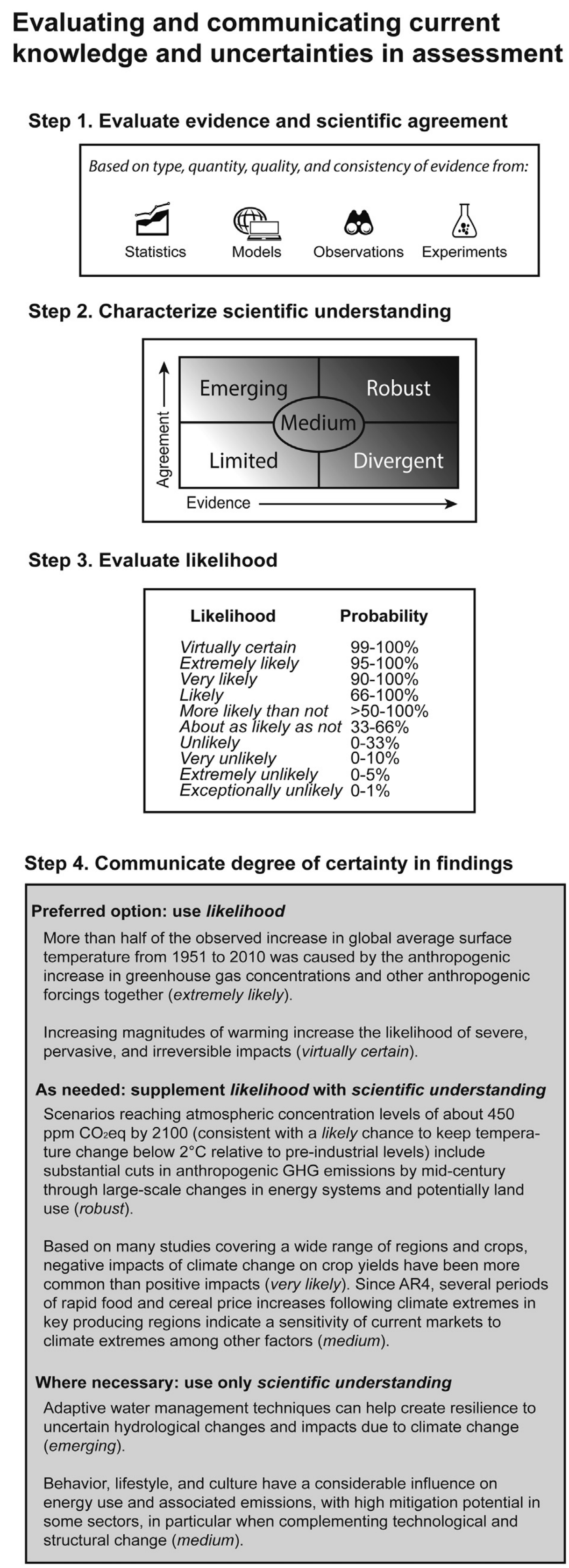

Fig. 7. (Continued)

5.2. Best practices for Integrating Evidence and Expert Judgment

Rigorous expert judgment is central to the added value of assessment. But effectively harnessing expert judgment requires deliberate attention to its role, tapping opportunities while preventing distortions. Deliberate practices can foster environments encouraging exploration of full ranges of perspectives on available evidence. They can also introduce tools and structured approaches for doing so.

In this section, we suggest options for enhancing the effectiveness of expert judgment in assessment, underpinning the development and communication of assessment conclusions (Section 5.1). We build from IPCC experiences (Sections 2–4), including missed opportunities. We integrate all of the associated experiences with broader recommendations from decision-support and assessing-assessments literature. As our starting point, the IPCC AR5 relied on expert judgments developed collectively. Author teams evaluated the literature, responded to abundant reviewer feedback, and collectively identified and revised assessment findings through iterative group deliberations and drafting rounds. In many cases, there was relatively little explicit attention to biases that can creep into expert judgment and group dynamics, and there were probably too many cases where personalities and norms, more than balanced attention to logic and ideas, steered conversations and judgments (O'Reilly et al., 2012; Brysse et al., 2013; Sutherland and Burgman, 2015). Despite calls over past years and guidance for authors (e.g., Section 2.2; Moss and Schneider, 2000; Yohe and Oppenheimer, 2011), assessment has not consistently and thoroughly incorporated advanced approaches for integrating evidence and expert judgment in the formulation of findings and the treatment of uncertainties. Moving into the future, assessment also has compelling opportunities for embedding more sophisticated approaches to individual versus group dynamics, tapping the best of both. Below we propose options that extend beyond the guidance described in Section 2.2. Associated practices would influence overall patterns of uncertainty characterizations (e.g., Section 3) as well as specific choices made in developing assessment findings (e.g., Section 4).

- **Train authors in expert judgment:** Starting assessment with author training can help minimize inevitable expert-judgment biases, including availability, anchoring, and overconfidence (Kahneman and Klein, 2009; Morgan, 2014). Training should prioritize the social and cultural dimensions of fostering rigorous expert judgment, as well as specific tools and mechanics for doing so. Training should be thorough especially for authors leading drafting teams and group deliberations. Additionally, assessment authors could be encouraged to try and experiment with any of the options below (Yohe and Oppenheimer, 2011). Training should create comfort with the fact that rigorous expert judgment may sometimes expand estimated ranges of values or outcomes. Training could be through written guidance, in-person sessions, or web-based meetings at the start of an assessment. Expert-judgment training could also be incorporated in broader assessment "boot camps," bringing authors up to speed on best practices for establishing environments of mutual respect in international collaborations, on literature synthesis techniques, on development of effective assessment graphics, on solicitation of decision-maker perspectives throughout the assessment, on clear and accessible communication, etc.
- **Improve group deliberations:** Group work can improve expert judgments and advice. At the same time, group interactions, especially when unstructured, can be subject to confirmation biases and groupthink. Potential pitfalls can be addressed in a variety of ways. For example, convening heterogeneous groups can help ensure consideration of full ranges of views and simultaneously ameliorate inevitable tendencies toward converging perspectives (Straus et al., 2009; Sutherland and Burgman, 2015; Corbera et al., 2016). And such groups can improve consideration of full ranges of hypotheses, perspectives, and outcomes through formal procedures (e.g., qualitative scenario techniques, what-if analyses, premortem explorations, team A/B analyses, red teaming), as well as informal approaches (e.g., presenting options simultaneously rather than sequentially, emphasizing consideration of alternatives, considering disconfirming or stretching questions, extending discussion time when needed) (Kahneman and Klein, 2009; Straus et al., 2009; Zenko, 2015; Trutnevyte et al., 2016). Trained group leaders or facilitators should focus on decision processes that foster open, interactive discussion and improve balanced documentation of ranges of views on supporting evidence. Benefits from structuring group interactions can be realized in both face-to-face and non-collocated deliberations (e.g., Dalal et al., 2011).
- **Rigorously synthesize literature:** Meta-analyses and systematic reviews (e.g., Poloczanska et al., 2013; Wittmann and Portner, 2013; Challinor et al., 2014) can provide foundational overviews of available evidence. In developing assessment, author teams should prioritize use of such approaches, for example using documented ranges and distributions in the literature to inform expert judgments about qualitative understanding. Use of formalized integration of expert judgment through elicitation should also be considered (Oppenheimer et al., 2016). More broadly, traceable accounts for conclusions should transparently indicate literature evaluated and related processes of expert judgment (IAC, 2010). To be effective, integration must fully value diverse forms of evidence, for example by not unduly prioritizing results from a particular class of models (e.g., Oppenheimer et al., 2007).
- **Foster awareness for expert judgments:** A focus on key factors and their influences, including through graphical depictions, can foster awareness of determinants of outcomes and associated uncertainties (Morgan, 2014). Straightforward approaches such as listing factors on cards and then sorting them by strength of influences and sources of uncertainty can further enhance considered judgments. These types of techniques can decrease tendencies to overweight shared but incomplete information in group discussions (Straus et al., 2009). They can help experts, individually and collectively, keep key factors in mind while organizing their thinking and judgments. Such techniques may also help counter tendencies to downplay important deep or structural uncertainties in consensus-based group deliberations and drafting (Oppenheimer et al., 2007).
- **Document individual views:** A large, yet often untapped opportunity for assessment is documenting individual expert perspectives that inform collective expert judgments. Doing so can enhance the entire process of expert judgment. For example,

Fig. 7. A simplified rigorous framework for characterizing degrees of certainty in assessment findings, building from the IPCC AR5 expert-judgment approach. This diagram illustrates a proposed process for evaluating and communicating knowledge and uncertainties in assessment. The process begins with evaluation of evidence and agreement, with scientific understanding characterized qualitatively (steps 1–2). Where possible, authors subsequently evaluate the likelihood of outcomes, values, states, or processes, based on subjective probabilistic judgments of all relevant evidence and uncertainties (step 3). Note that the likelihood categories should be considered to have "fuzzy" boundaries. Authors present scientific-understanding or likelihood terms with assessment conclusions, communicating their expert judgments accordingly (step 4). Icons in step 1 were made by Freepik and Sarfraz Shoukat from www.flaticon.com. Examples in step 4 are adapted from IPCC AR5 SPM conclusions across working groups.

individual surveys (e.g., Gattuso et al., 2013) or individually elicited probability distributions (e.g., Kriegler et al., 2009; Zickfeld et al., 2010) can be used to document perspectives as input for group deliberations on available evidence (CCSP, 2009; Morgan, 2014). Documented quantitative beliefs (e.g., about outcome probabilities, values of key parameters, or likelihoods of different causal processes) can provide an important foundation for more holistic collective judgments about the state of knowledge, for both well-established and more speculative understanding. Such documentations should form part of an author team's traceable account for their conclusions, especially for likelihood assignments (Section 5.1). They could reveal the diversity of expert views, hypotheses, and mental models, as well as their evolutions through the course of the assessment process. In this way, individually documented perspectives would increase transparency of expert judgments underpinning assessment conclusions, especially compared to qualitative summaries that can mask disagreements (Oppenheimer et al., 2007; IAC, 2010; Socolow, 2011; Anderegg et al., 2014). They could also decrease tendencies toward overconfidence and enhance evaluations of full ranges of outcomes, including more extreme, low-probability/high-consequence outcomes. Such methods should be attentive to fundamental features of topics of interest. For example, issues may range from matters of fact dependent on empirical science, to realms where individual and social behavior will determine outcomes in much less predictable ways.

- **Acknowledge and explore normative dimensions of assessment**: Normative judgments inevitably affect the assessment process, for example through different worldviews, value-based perspectives, and subjective assumptions infused across disciplines (CCSP, 2009; Schneider, 2010; Yohe and Oppenheimer, 2011; Adler and Hadorn, 2014; Victor, 2015). But such normative dimensions have often been implicit, rather than explicit, in assessment, sometimes constraining or confounding the assessment of risks and response options. For example, different emphases can be placed on impacts fully versus inadequately captured in monetary terms or on impacts for the present versus the future; such distinct emphases involve value-based dimensions shaping assessment of what is at stake, even though associated subjective aspects may not be explicitly acknowledged or addressed. Opinion surveys, evaluation and ranking of subjective preferences, or development of alternative conclusions across explicit value-based framings could help document subjective dimensions of assessment directly, increasing the nuance, transparency, and effectiveness of resulting conclusions. Additionally, conclusions can inform value- and objective-based decisions by clarifying dimensions contingent on human choice and intention as compared to other uncertainties (Swart et al., 2009).

Implementing such practices would likely benefit from a two-tiered approach, spearheaded by the overall leadership and staff facilitating an assessment. First, authors in leading roles (e.g., coordinating lead authors responsible for each IPCC chapter) could receive in-depth, in-person training at the very start of an assessment. These experts have the greatest ability to shape approaches and atmospheres for expert judgment, as well as the assessment that results. Through specific tips, tools, and simulations, such training could thoroughly equip assessment leaders to implement best practices across the above domains. Effective training could particularly advance group deliberations that improve rather than hinder expert judgments (see "Improve group deliberations" above). It could also enhance understanding of methods for considering uncertainty sources and encompassing them effectively in resulting assessment conclusions (see "Foster awareness for expert judgments" above).

Second, before and throughout the assessment, all authors could receive guidance on implementing best practices through short written summaries, web-based tutorials, and meeting presentations. In the IPCC AR5, for example, authors received extended guidance on the expert-judgment framework (Section 2.1) but relatively less how-to guidance on associated best practices for expert judgment more broadly (Section 2.2). Step-by-step guidance could productively address options for synthesizing literature, approaches for documenting key factors and uncertainties, methods for documenting individual views in chapter traceable accounts, and ways to acknowledge and transparently incorporate normative dimensions of assessment (Section 5.2). Step-by-step guidance could also address lessons unfolded in Section 5.3. In addition to initial orientations, all authors could benefit from ongoing feedback, debriefs, and exchanges throughout the assessment, advancing learning-by-doing and whole-report consistency.

Implementing best practices would require substantial investment by the leaders and staff facilitating an assessment. Thoughtful complaint-resolution processes could complement the two-tiered approach above. They could ensure group dynamics get back on track when problems arise and actively promote environments of mutual respect for rigorous expert judgment. With commitment from all involved, a thorough investment toward best practices would likely yield returns, maximizing the value of experts' time dedicated to the effort and potentially driving integrative assessment to new peaks of relevance and impact.

5.3. Accessible, Clear Communication of Expert Judgment

Even when expert judgments are carefully developed in assessment (Section 5.2), rigorously and clearly communicating them has some inherent tensions. First, transparency of expert judgments can reduce the accessibility of assessment findings, for example where it introduces cumbersome wording. Second, the simplest expert-judgment frameworks, even though easiest to understand for the reader, may fail to encompass assessment needs across disciplines (e.g., Adler and Hadorn, 2014). Third, intuitive interpretations of expert-judgment terms may stray from their rigorous definitions (e.g., Budescu et al., 2014). Overall, assessment conclusions must be scientifically rigorous, integrative, nuanced, unambiguous, relevant, and compelling, all while using words non-experts can understand and syntax that is easy to read. Effective communication in assessment is fundamentally important and also difficult.

Based on the IPCC AR5's expert-judgment experiences, we learned several lessons for accessible communication of expert judgment in assessment (Section 5.1):

- **Present expert judgments clearly:** Assessment conclusions should be worded to crisply articulate key findings. Allowing degree-of-certainty terms to dominate the reading experience greatly reduces accessibility. Wherever possible, terms should be unambiguous but not cumbersome, for example through parenthetical presentation at the end of a sentence rather than in-line presentation (e.g., avoiding, "There is *medium confidence* that it is *very likely* that...").
- **Present expert judgments transparently:** Accessible presentation of degree-of-certainty terms should be accompanied by more complete and rigorous indication of underlying expert judgments, including the range of expert views (e.g., Oppenheimer et al., 2016). In particular, expert judgments underpinning conclusions should be transparent. Report readers should be able to readily connect high-level conclusions to traceable accounts of underlying evidence and expert judgments. Presenting individ-

ual, not just collective, subjective probabilistic judgments can enhance communication of low-probability/high-consequence outcomes relevant to risk management and decision-making. Being effectively transparent may often involve scaffolding: presenting easy-to-understand high-level conclusions with clear orientation on where readers can find complete explanation of judgments entailed.

- **Communicate conclusions with nuance:** Developing nuanced conclusions is important. Appropriately generalizing understanding is essential, avoiding weakly supported overgeneralizations (Section 2.2). High-confidence truisms or likelihoods for ill-defined outcomes may do little to inform decisions (IAC, 2010). In addition, sophisticated unfolding of lower-certainty findings (Section 4) can help facilitate productive interpretations.
- **Understand the audience:** Actively soliciting decision-maker inputs through reviews, surveys, interviews, and structured discussions increases understanding of relevant questions for assessment (e.g., Pidgeon and Fischhoff, 2011). Sustaining engagement with communications and decision-support specialists can maximize the likelihood that conclusions will resonate with intended audiences. Further, evaluating and refining messages with target audiences is essential for effective communication (CCSP, 2009). Short orienting overviews, briefing notes, graphics and animations, interactive tutorials, and videos can all enhance the accessibility of assessment findings. Many of these priorities have been increasingly explored in IPCC assessments. Further engaging stakeholders in the assessment process could increase its effectiveness.
- **Ensure accurate interpretations of degree-of-certainty terms:** Indicating probability ranges for different likelihood terms is essential because their definitions do not necessarily match intuitive interpretations (Patt and Dessai, 2005; Budescu et al., 2014). In addition to initial introduction, continued supplementing with numerical probability ranges can better align reader interpretations with intended meanings. Providing example rankings of probabilities of different risks can also help make probability concrete rather than abstract (Patt and Dessai, 2005).

5.4. Concluding Thoughts on Expert Judgment in Assessment

The IPCC AR5 is the most comprehensive environmental assessment to date. It represents an unprecedented organization and communication of expert judgments about a vast scientific literature across many chapters and policymaker summaries. These extensive experiences in harnessing expert judgment also provide lessons for the future. We have evaluated both the advances and challenges in expert-judgment approaches across the IPCC AR5. Our proposed improvements and practices offer compelling opportunities for evaluating and communicating the state of knowledge in future assessments. Advancing expert judgment will further robust scientific foundations on the many topics underlying decisions and actions toward a sustainable future.

Acknowledgments

We thank J. Mach for research assistance. Funding for this work was provided by the Alexander von Humboldt Foundation.

References

Adler, C.E., Hadorn, G.H., 2014. The IPCC and treatment of uncertainties: topics and sources of dissensus. WIREs Climate Change 5, 663–676.
Anderegg, W.R.L., Callaway, E.S., Boykoff, M.T., Yohe, G., Root, T.L., 2014. Awareness of both Type 1 and 2 errors in climate science and assessment. Bulletin of the American Meteorological Society 95, 1445–1451.
Aven, T., Renn, O., 2015. An evaluation of the treatment of risk and uncertainties in the IPCC reports on climate change. Risk Analysis 35, 701–712.
Brysse, K., Oreskes, N., O'Reilly, J., Oppenheimer, M., 2013. Climate change prediction: Erring on the side of least drama? Global Environmental Change 23, 327–337.
Budescu, D.V., Por, H.-H., Broomell, S.B., Smithson, M., 2014. The interpretation of IPCC probabilistic statements around the world. Nature Climate Change 4, 508–512.
CCSP, 2009. Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Decisionmaking. United States Climate Change Science Program, National Oceanic and Atmospheric Administration, Washington, DC.
Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change 4, 287–291.
Corbera, E., Calvet-Mir, L., Hughes, H., Paterson, M., 2016. Patterns of authorship in the IPCC Working Group III report. Nature Climate Change 6, 94–99.
Dalal, S., Khodyakov, D., Srinivasan, R., Straus, S., Adams, J., 2011. ExpertLens: A system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge. Technological Forecasting and Social Change 78, 1426–1444.
Drouet, L., Bosetti, V., Tavoni, M., 2015. Selection of climate policies under the uncertainties in the Fifth Assessment Report of the IPCC. Nature Climate Change 5, 937–940.
Gattuso, J.-P., Mach, K.J., Morgan, G., 2013. Ocean acidification and its impacts: An expert survey. Climatic Change 117, 725–738.
GEA, 2012. Global Energy Assessment: Toward a Sustainable Future. Cambridge University Press, Cambridge, United Kingdom and the International Institute for Applied Systems Analysis, Laxenburg, Austria.
IAC, 2010. Climate Change Assessments: Review of the Processes and Procedures of the IPCC. InterAcademy Council, Amsterdam, The Netherlands.
IPBES, 2016. In: Potts, S.G., Imperatriz-Fonseca, V.L., Ngo, H.T., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J., Vanbergen, A.J., Aizen, M.A., Cunningham, S.A., Eardley, C., Freitas, B.M., Gallai, N., Kevan, P.G., Kovács-Hostyánszki, Kwapong, P.K., Li, J., Li, X., Martins, D.J., Nates-Parra, G., Pettis, J.S., Viana, B.F. (Eds.), Summary for Policymakers of the Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on Pollinators, Pollination and Food Production. pp. 1–28., http://www.ipbes.net/sites/default/files/downloads/Pollination_Summary%20for%20policymakers_EN_.pdf.
IPCC, 2005. Guidance notes for lead authors of the IPCC Fourth Assessment Report on addressing uncertainties. Intergovernmental Panel on Climate Change, Geneva, Switzerland.
IPCC, 2007a. Summary for Policymakers. In: Solomon, S., Qin, D., Manny, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 1–18.
IPCC, 2007b. Summary for Policymakers. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E. (Eds.), Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 7–22.
IPCC, 2007c. Summary for Policymakers. In: Metz, B., Davidson, O.R., Bosch, P.R., Dave, R., Meyer, L.A. (Eds.), Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report on the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 1–23.
IPCC, 2007d. Summary for Policymakers. In: Pachauri, R.K., Reisinger, A. (Eds.), Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland, pp. 1–22.
IPCC, 2013a. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, p. 1535.
IPCC, 2013b. Summary for Policymakers. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 3–29.
IPCC, 2014a. In: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, p. 1132.
IPCC, 2014b. In: Barros, V.R., Field, C.B., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment

Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, p. 688.

IPCC, 2014c. In: Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickermeier, P., Kriemann, B., Savolainen, J., Schlomer, S., von Stechow, C., Swickel, T., Minx, J.C. (Eds.), Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, p. 1435.

IPCC, 2014d. In: Pachauri, R.K., Meyer, L.A., Core Writing Team, (Eds.), Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland, p. 151.

IPCC, 2014e. Summary for Policymakers. In: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 1–32.

IPCC, 2014f. Summary for Policymakers. In: Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickermeier, P., Kriemann, B., Savolainen, J., Schlomer, S., von Stechow, C., Swickel, T., Minx, J.C. (Eds.), Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 1–30.

IPCC, 2014g. Summary for Policymakers. In: Pachauri, R.K., Meyer, L.A., Core Writing Team, (Eds.), Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland, pp. 1–31.

Kahneman, D., Klein, G., 2009. Conditions for intuitive expertise: A failure to disagree. American Psychologist 64, 515–526.

Karl, T.R., Melillo, J.M., Peterson, T.C. (Eds.), 2009. Global Climate Change Impacts in the United States. Cambridge University Press, New York, USA.

Kriegler, E., Hall, J.W., Held, H., Dawson, R., Schellnhuber, H.J., 2009. Imprecise probability assessment of tipping points in the climate system. Proceedings of the National Academy of Sciences 106, 5041–5046.

Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., 2010. Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties, https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf.

Mastrandrea, M.D., Mach, K.J., 2011. Treatment of uncertainties in IPCC assessment reports: Past approaches and considerations for the Fifth Assessment Report. Climatic Change 108, 659–673.

Mastrandrea, M.D., Mach, K.J., Plattner, G.-K., Edenhofer, O., Stocker, T.F., Field, C.B., Ebi, K.L., Matschoss, P.R., 2011. The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. Climatic Change 108, 675–691.

McDonald, J.H., 2014. Handbook of Biological Statistics, 3rd Ed. Sparky House Publishing, Baltimore, USA.

MEA, 2005. In: Hassan, R., Scholes, R., Ash, N. (Eds.), Ecosystems and Human Well-Being: Current State and Trends. Millennium Ecosystem Assessment. Island Press, Washington, DC, p. 917.

Melillo, J.M., Richmond, T.T.C., Yohe, G.W. (Eds.), 2014. Climate Change Impacts in the United States: The Third National Climate Assessment. US Global Change Research Program, Washington, DC.

Mitchell, R.B., Clark, W.C., Cash, D.W., Dickson, N.M. (Eds.), 2006. Global Environmental Assessments: Information and Influence. MIT Press, Cambridge, USA.

Morgan, M.G., 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences 111, 7176–7184.

Moss, R.H., Schneider, S.H., 2000. Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In: Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC. World Meteorological Organization, Geneva, Switzerland, pp. 33–51.

NEAA, 2010. Assessing an IPCC assessment: An analysis of statements on projected regional impacts in the 2007 report. PBL Netherlands Environmental Assessment Agency. The Hague, The Netherlands.

O'Reilly, J., Oreskes, N., Oppenheimer, M., 2012. The rapid disintegration of projections: The West Antarctic Ice Sheet and the Intergovernmental Panel on Climate Change. Social Studies of Science 42, 709–731.

Oppenheimer, M., O'Neill, B.C., Webster, M., Agrawala, S., 2007. The limits of consensus. Science 317, 1505–1506.

Oppenheimer, M., Little, C.M., Cooke, R.M., 2016. Expert judgement and uncertainty quantification for climate change. Nature Climate Change 6, 445–451.

Patt, A., Dessai, S., 2005. Communicating uncertainty: lessons learned and suggestions for climate change assessment. Comptes Rendus Geoscience 337, 425–441.

Pidgeon, N., Fischhoff, B., 2011. The role of social and decision sciences in communicating uncertain climate risks. Nature Climate Change 1, 35–41.

Poloczanska, E.S., Brown, C.J., Sydeman, W.J., Kiessling, W., Schoeman, D.S., Moore, P.J., Brander, K., Bruno, J.F., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Holding, J., Kappel, C.V., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Schwing, F., Thompson, S.A., Richardson, A.J., 2013. Global imprint of climate change on marine life. Nature Climate Change 3, 919–925.

Schneider, S.H., 2010. A constructive deconstruction of deconstructionists: A response to Demeritt. Annals of the Association of American Geographers 91, 338–344.

Socolow, R.H., 2011. High-consequence outcomes and internal disagreements: Tell us more, please. Climatic Change 108, 775–790.

Straus, S.G., Parker, A.M., Bruces, J.B., Dembosky, J.W., 2009. The group matters: A review of the effects of group interaction on processes and outcomes in analytic teams. RAND Corporation, Santa Monica, California.

Sutherland, W.J., Burgman, M., 2015. Policy advice: Use experts wisely. Nature 526, 317–318.

Swart, R., Bernstein, L., Ha-Duong, M., Petersen, A., 2009. Agreeing to disagree: uncertainty management in assessing climate change, impacts and responses by the IPCC. Climatic Change 92, 1–29.

Trutnevyte, E., Guivarch, C., Lempert, R., Strachan, N., 2016. Reinvigorating the scenario technique to expand uncertainty consideration. Climatic Change 135, 373–379.

Victor, D.G., 2015. Embed the social sciences in climate policy. Nature 520, 27–29.

Wittmann, A.C., Portner, H.-O., 2013. Sensitivities of extant animal taxa to ocean acidification. Nature Climate Change 3, 995–1001.

WMO, 2014. Assessment for Decision-Makers: Scientific Assessment of Ozone Depletion: 2014. Global Ozone Research and Monitoring Project—Report No. 56. World Meteorological Organization, Geneva, Switzerland.

Yohe, G., Oppenheimer, M., 2011. Evaluation, characterization, and communication of uncertainty by the intergovernmental panel on climate change—an introductory essay. Climatic Change 108, 629–639.

Zenko, M., 2015. Red Team: How to Succeed by Thinking Like the Enemy. Basic Books, New York, USA.

Zickfeld, K., Morgan, M.G., Frame, D.J., Keith, D.W., 2010. Expert judgments about transient climate response to alternative future trajectories of radiative forcing. Proceedings of the National Academy of Sciences 107, 12451–12456.