High-consequence outcomes and internal disagreements: tell us more, please

Robert H. Socolow

Abstract This article is one of several in this special double-issue that reports the views of "users" of IPCC reports. I am a user in the sense that I advise the policy-making community and rely on the IPCC reports to provide me with authoritative views on the state of the science. My principal recommendation for making the IPCC more helpful to the policy-making community is to strive in the Fifth Assessment Report (AR5) to communicate fully what the climate science community understands and does not understand about high-consequence outcomes. This will require the AR5 authors to provide vivid information about future worlds where high-consequence outcomes have emerged. It will also require the AR5 authors to reveal any disagreements persisting among them after the give-and-take of the writing process has run its course. In the Fourth Assessment Report (AR4) the presentation of high-consequence outcomes had shortcomings that can be rectified in AR5.

1 Helping the policymaker decide how important climate change is

Users like me, who read IPCC reports in order to give better advice to others, are probably a small fraction of all the users. Many of us, including me, are self-appointed. We publish (including articles like this one), lecture, and serve on committees. In our own minds, we do not "represent" any interest group, and our reward for doing well is to be invited to serve again. Our principal target is the policy-making community, including office holders, their staffs, and those opinion-formers whom they take seriously. In giving advice, we hunt for the issues that matter a lot but are poorly formulated or are buried out of sight. The better the IPCC does its work, the less work we have.

Inevitably, those with a role like mine make assumptions about what policymakers want to know. I assume that Objective Number One for anyone in the policy-making community is to form an independent judgment of how urgent the climate change problem is. Does the world need to drop everything else and assign climate change the highest priority, or is

R. H. Socolow (✉)
Department of Mechanical and Aerospace Engineering, Princeton University, Guyot 139,
Princeton, NJ 08544, USA
e-mail: socolow@princeton.edu

climate change one of several important problems? Does the world have the luxury of being able to take a whole century to accomplish decarbonization, or must the job be accomplished in just two or three decades?

Policymakers understand that a fast pace for emissions reduction will generate large disruptions as "solutions" are implemented, but that a fast pace will also produce large benefits to the extent that disruptions from climate change itself are avoided. Thus, to make an informed decision about pace, a policymaker needs help answering two questions: How disruptive are the "solutions"? How disruptive is climate change?

Disruptions associated with solutions, and ways for the IPCC to discuss them, are topics for a different essay. Suffice it to say that policymakers prefer to go slow. They are inured to the sales pitches they receive from all quarters. In considering extensive reshaping of landscapes, global expansion of nuclear power, albedo management, and other problematic responses to climate change, they can imagine cures worse than the disease. In my view, the Fifth Assessment Report (AR5) could distinguish itself, relative to its predecessors, by the depth of its probing of the problematic aspects of "solutions."

But the objective of *this* essay is to address how the IPCC could better help the policy-making community in its search for deeper understanding of disruptions due to climate change itself. I set aside the well known conundrum that the policy maker seeks spatial detail (especially related to the region he or she represents) while the modeler is constrained by coarse resolution. Rather, my focus here is on the policy maker's interest in understanding *high-consequence outcomes*.

The phrase "high-consequence outcomes" has become a term of art in IPCC literature. High-consequence outcomes arise in a world of very strong positive feedbacks, a world that spirals out of control. For specificity, I have in mind a five-meter rise in sea level by the end of this century, or major alterations of the global hydrocycle, or major changes in ocean biota, or yet another of what my colleague, Steve Pacala, calls "monsters behind the door." The key recommendation of this article is this: *In AR5 the IPCC should communicate fully what the science community does and does not understand about high consequence outcomes.*

The rest of this article is intended to shed light on what this task entails.

2 Making high-consequence outcomes vivid

The policy-making community understands that there are two complementary ways to think about damage from climate change. There are outcomes that are best guesses, and there are outcomes that are either much more benign or much more alarming than the best guesses. Policymakers know they need to understand the full story, the full distribution—both the middle and the "tails." They wonder whether, in reaching a judgment about whether to act forcefully now, they should give more weight to the best guesses or to the tails, and they want to understand the relative weights.

Alas, earth systems science right now cannot provide much guidance about these relative weights, yet they matter greatly. Martin Weitzman has shown that the problem of relative weights for mean vs. extreme outcomes can be formulated mathematically using a damage function that sums over outcomes, and that whether the mean or the extreme dominates (whether the sum converges) depends on currently unknowable details of the probability distribution for the worst outcomes. [Weitzman 2009] Accordingly, the policy-making community needs information about both probable and improbable outcomes. One can imagine that, for many policymakers, the priority given to climate change is strongly dependent on what the IPCC thinks about high-consequence outcomes.

Of course, low-consequence outcomes are also important. The policymaker wants to know if there could be credible, relatively benign outcomes, or, more accurately, outcomes that remain benign for a relatively long time. A complete report on the distribution of outcomes would discuss these outcomes as well.

Indeed, AR5 will lead to better public discourse if it helps the policymaker understand that disparate outcomes are consistent with current earth-systems science. Imagine a museum exhibit based on AR5 that, in two adjacent displays, presents two worlds at some future date with the same greenhouse gas concentrations: in one display, we have been lucky and the manifestations of climate change have turned out to be relatively gentle, and in the other we have been unlucky and they have turned out to be severe. Perhaps, there would be *two* pairs of displays, one pair relevant to a much higher greenhouse concentration than today's, and the other pair relevant to a moderately higher concentration. Perhaps the display corresponding to being lucky at the high concentration would be identical to the display for being unlucky at the moderate concentration.

Producing such displays is within the capabilities of the climate science community. Each computer model of the earth system produces a distribution of responses to changes in greenhouse gas concentrations, as key variables are varied within a range. Outcomes that deviate substantially from the mean outcomes—in both directions—are among the outputs of these exercises. If AR5 can report a few of these deviant outcomes in detail (perhaps, in pairs), the public will be enlightened and the IPCC will not make itself vulnerable to charges of scaremongering.

The late Sidney Coleman, a theoretical physicist, followed the dictum: "In order to know the truth, it is necessary to imagine a thousand falsehoods." In understanding climate change, models help us do the imagining, but only if there is a general sharing of provocative runs of models before these runs are lost in an averaging process. As a Coleman disciple, I restated this thought many years ago: "Analyses are not what people care about…Disciplined analysis has enormous unused capability…I look forward to the day when it is usual to have more open, more self-critical, even more playful discourse." [Socolow 1976, pp. 1, 2, 32]

The IPCC has already internalized that it must deal with high-consequence outcomes in AR5. Its job is to be responsive to the needs of the participating governments (nearly all the governments of the world). These needs were expressed on the concluding page of the Summary for Policymakers in the Synthesis Report of the Fourth Assessment Report (AR4), which observes that "responding to climate change involves an iterative risk management process." [IPCC 2007a, p. 22] Three sentences later, "potentially catastrophic losses" are invoked as one of the contributions to the damage assessment that is integral to risk management. By implication, for AR5 to assist governments in conducting iterative risk management, it must characterize all significant risks—including those that arise from high-consequence outcomes, even those with low probability. [Gary Yohe, private communication.]

With the best will in the world, however, experience suggests that the authors of AR5 will find it difficult to communicate what the science community does and doesn't understand about high-consequence outcomes. Some of the difficulties that can be expected are revealed in the story of how AR4 dealt with the equilibrium climate sensitivity (ECS). The ECS, writes the IPCC, is "the equilibrium annual global mean temperature response to a doubling of equivalent atmospheric CO_2 from pre-industrial levels and is thus a measure of the strength of the climate system's eventual response to greenhouse gas forcing." [IPCC 2007b, p. 718.] ECS is the principal variable long chosen by climate scientists to communicate outcome-related uncertainties to the public.

3 AR4 and the equilibrium climate sensitivity

I became interested in the treatment of ECS uncertainty in AR4 in the summer of 2010 as I was drafting a segment of the report of the Committee on America's Climate Choices (CACC) of the National Research Council. [National Research Council 2011] To address uncertainty in climate outcomes, I decided to build on the discussion of ECS in AR4. I knew that AR4, like the earlier Assessment Reports, summarizes its findings by presenting a range for ECS, and that the AR4 range extends from 2.0°C to 4.5°C. I also knew that in AR4, for the first time, a probability is attached to the ECS range. In AR4, the chosen range from 2.0°C to 4.5°C is called "likely," and "likely" is defined in many places in the IPCC report to mean "having 66% probability."

I was especially interested in using this probability to introduce the layperson to climate feedbacks. I could subtract 66% from 100%, divide by two, and report that AR4 associates 17% probability with an ECS above 4.5°C, as well as with an ECS below 2.0°C. Rounding off, I could report that the world's climate scientists had arrived at a consensus that there is one chance in six that the ECS exceeds 4.5°C. The CACC as a whole could then write the paragraphs that would fill in the picture of what a 4.5°C rise in the long-term average surface temperature might entail.

However, nowhere in AR4 does it say that there is a 17% probability that the equilibrium climate sensitivity exceeds 4.5°C! Instead, AR4 reports that values of ECS above 4.5°C "cannot be excluded." That certain values of a variable "cannot be excluded" implies, to me at least, that an exclusion process is under way and that in due course these high values *will* be excluded. "Cannot be excluded" sounds lawyerly—the sort of language one finds in a stock prospectus to describe risks a company doesn't actually consider real. By contrast, 17% focuses the mind: 17% is the probability of death in one round of Russian roulette. I decided to embark on a reconnaissance to find out why the IPCC chose "cannot be excluded" over "17%." One reason I am writing this article is to share what I learned.

In response to my queries, I was provided with Fig. 1. The Figure does not appear in any IPCC report. It was produced by Martin Manning, one of the leaders of AR4's WG1, to assist WG1 in its deliberations. It shows a large blank region at the right of the Figure, characterized by both "17%" and "Cannot be excluded," the choice between them still not made. In his words in a note to me:

> I put that graph together to see if the LAs [Lead Authors] would agree to some way of bringing all the percentage probabilities for the ranges together consistently…It would have been confusing if the implied upper bounds on probabilities for each of the four different sub-ranges would have added up to more than 100%.

Although Manning did not provide me with the Figure, he has given me permission to publish it.

Figure 1 provides a window on current climate science. Any long-term stabilized value of the aggregate concentration of greenhouse gases (measured as a "CO_2-equivalent" concentration, in ppm) produces an average temperature for the earth's surface. A horizontal line segment for any specific concentration, extending from the left boundary of the dark blue region to its right boundary, identifies the range of values of this temperature that has a 66% probability of being realized. That same range is also called "likely." To the left and right of the dark blue region lie what are sometimes called the "low tail" and the "high tail" of the temperature distribution. Figure 1 assigns some structure for the low tail by presenting an additional curve that separates the lowest 10% of the probability distribution from the next 7%. No additional structure is shown for the high tail.

Figure 1 includes the distribution of the equilibrium climate sensitivity as a special case. The ECS, by definition, is the temperature on the x-axis that corresponds to a concentration
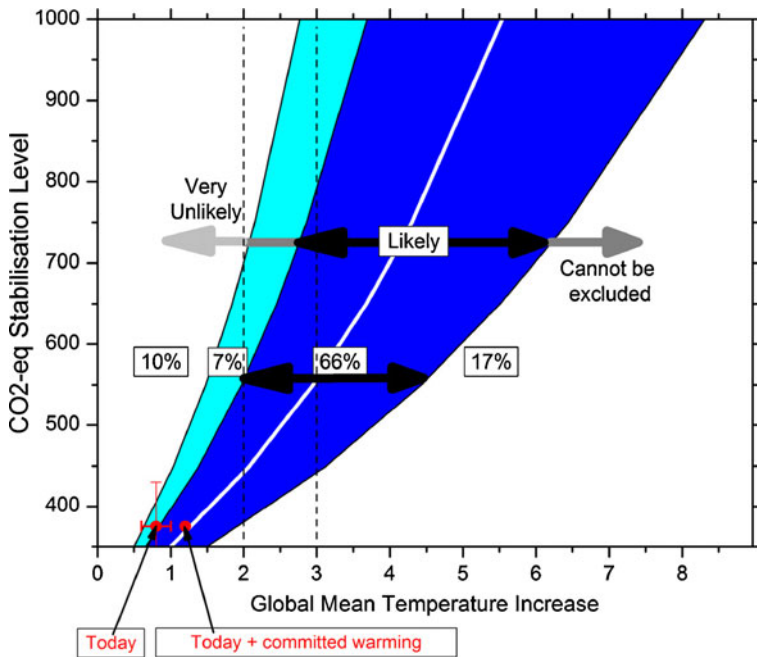
Fig. 1 The uncertain relationship between the stabilized concentration of greenhouse gases in the atmosphere and the long-term global mean increase in surface temperature (relative to pre-industrial times). The units for the x and y axes are °C and ppm, respectively. The Figure was prepared by Martin Manning to facilitate discussion in Working Group 1 during the preparation of the IPCC Fourth Assessment Report. The Figure was not published

on the y-axis that is double the pre-industrial concentration, or 560 ppm. The lower of the two dark, thick double-arrows shows the "likely" range of ECS, running from 2.0°C to 4.5°C.[1] Thus, a value of ECS above 4.5°C both has 17% probability and "cannot be excluded."

In September, 2010, Manning explained to me why he prepared Fig. 1:

[Evaluating climate sensitivity] "was quite a tricky issue for the WG1 AR4…Let me try and summarize… [We hoped to provide] a more definitive statement about the climate sensitivity than in all earlier IPCC assessments…It became an issue that brought together LAs [lead authors] from chapters 8, 9 and 10…But to move from all of those different studies to converge on a range required expert judgment on how to filter things while also covering different perspectives coming from recent and paleo data.

"So there was a lot of discussion about the constraints on ranges and this led to agreement across the authors on some things but not on a fully consistent way of putting together a complete probability distribution for equilibrium climate sensitivity (ECS)…There is a bit of a split between different experts in this area and in particular about putting a likelihood on ECS being >4.5 C. Some felt that this was 'very unlikely' while others argued that, because climate sensitivity is a very long term issue, that makes it very hard to pin down probability at the upper end of the range."

---

[1] The vertical dashed line at 2°C in Fig. 1 is another line with high policy relevance: 400 ppm and 550 ppm are the approximate bounds within which an ESC of 2°C is "likely" to be achieved.

Evidently, the experts weren't in agreement about how to assign adjectives and percents to Fig. 1. In effect, some experts wanted to label the dark blue region "very likely," instead of "likely," and to replace "66%" by "90%." They would then have assigned half of the rest to the white space at the right, or only 5%. These experts didn't prevail in the labeling of the dark blue region, but their views were reflected in assigning "cannot be excluded," rather than "17%," to the severe-outcome white region at the right.[2]

Note that the disagreement here has little to do with whether tails are "fat" or not. Discussions of fatness are about the details of the distribution *above* the upper boundary of "likely." How quickly the probability distribution falls at the highest temperatures determines, for example, whether the total damage function converges or diverges.

The AR4 WG1 report does devote considerable attention to disparate views about the equilibrium climate sensitivity held at the time. A key figure, "Comparisons between different estimates of the pdf (or relative likelihood) for ECS," superimposes nine probability density functions (pdf's) for ECS. [IPCC 2007b, Section 9.6, Figure 9.20, p. 720] The accompanying text explains the various methodologies used in these nine sources. The figure is reproduced below as Fig. 2.

The upper part of Fig. 2 shows the full probability density functions; the peak in the pdf for "Forest 06," for example, lies just above 2°C. The lower part of Fig. 2 shows summary statistics for each pdf shown in the upper part: the ends of each line segment show where 5% and 95% of the cumulative distribution are located, and the central dot shows where half of the distribution lies on either side. The 95% values in all but one case are well above 4.5°C; the 95% point for the "Forest 06" distribution, for example, is at almost 9°C The text accompanying this Figure in AR4 explains that among the reasons for the analysts' choice of such high 95% values is that "a high aerosol forcing could nearly cancel greenhouse gas forcing over the 20th century." [IPCC 2007b, p. 726]

Section 9.6 concludes with the observation: "Although upper limits can be obtained by combining multiple lines of evidence, remaining uncertainties that are not accounted for in individual estimates (such as structural model uncertainties) and possible dependencies between individual lines of evidence make the upper 95% limit of ECS uncertain at present." [IPCC 2007b, p. 727, one sentence from the end of Section 9.6] What a tortured sentence! The AR5 authors must do better. They too will confront disparate views of high-consequence events in individual publications. To serve their users, they will want to distill these views into accessible paragraphs that reveal the principal reasons for individual and collective uncertainty. In the next section, we suggest how this might be done.

4 Communicating internal disagreements in AR5

The core statement in AR4 about high consequence outcomes—that values of ECS greater than 4.5°C "cannot be excluded"—was the outcome of a search for consensus. Scholars of group processes know a lot about consensus-building. Each of the individuals engaged in building a consensus about the value of some parameter can be expected to be overconfident about how close his or her own numerical estimate is to the "real" value

---

[2] On the basis of Fig. 1 one can speculate that there was little disagreement about assigning 17% probability to ECS values below 2°C. "Cannot be excluded" may have emerged as the preferred alternative for those who would otherwise have had to advocate for reporting an asymmetric range (having 17% probability below its lower boundary and 5% probability above its upper boundary).
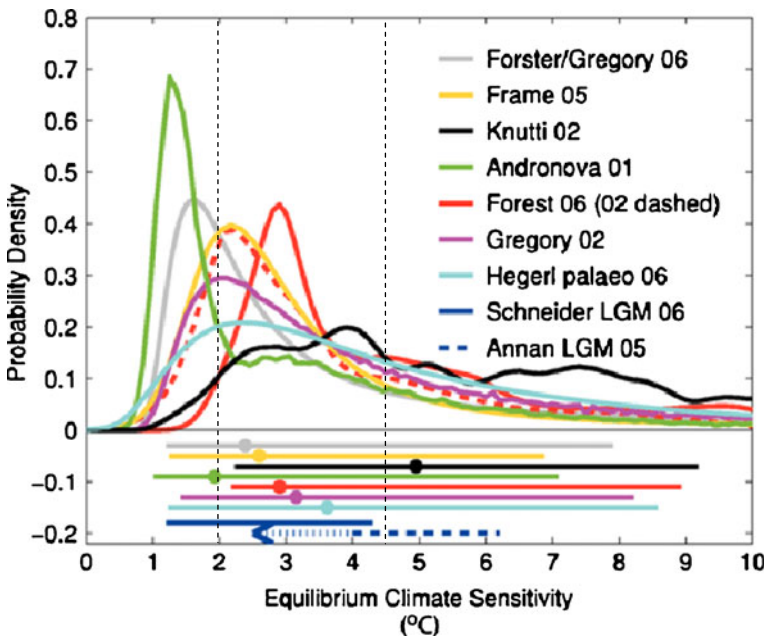
Fig. 2 Identical to Figure 9.20 in [IPCC 2007b, p. 720], except for the addition of vertical lines at 2°C and 4.5°C to assist the reader. *Partial caption*: "Comparison between different estimates of the PDF (or relative likelihood) for ECS (°C). All PDFs/likelihoods have been scaled to integrate to unity between 0°C and 10°C ECS. The bars show the respective 5 to 95% ranges, dots the median estimate…." *Note*: Probability densities can only be positive or zero: the two negative y-axis values (-0.1 and -0.2) are meaningless and should have been removed. (The full caption is: "Figure 9.20. Comparison between different estimates of the PDF (or relative likelihood) for ECS (°C). All PDFs/likelihoods have been scaled to integrate to unity between 0°C and 10°C ECS. The bars show the respective 5 to 95% ranges, dots the median estimate. The PDFs/likelihoods based on instrumental data are from Andronova and Schlesinger (2001), Forest et al. (2002; dashed line, considering anthropogenic forcings only), Forest et al. (2006; solid, anthropogenic and natural forcings), Gregory et al. (2002a), Knutti et al. (2002), Frame et al. (2005), and Forster and Gregory (2006), transformed to a uniform prior distribution in ECS using the method after Frame et al. (2005). Hegerl et al. (2006a) is based on multiple palaeoclimatic reconstructions of NH mean temperatures over the last 700 years. Also shown are the 5 to 95% approximate ranges for two estimates from the LGM (dashed, Annan et al., 2005; solid, Schneider von Deimling et al., 2006) which are based on models with different structural properties. Note that ranges extending beyond the published range in Annan et al. (2005), and beyond that sampled by the climate model used there, are indicated by dots and an arrow, since Annan et al. only provide an upper limit. For details of the likelihood estimates, see Table 9.3. After Hegerl et al. (2006a)."

that will eventually emerge; each underestimates the true range of uncertainty. [Morgan and Keith 2008] When they get together as a group and confront each other's narrow estimate, one might imagine that their search for consensus would lead to the reporting of broad uncertainty. However, this often does not happen. Oppenheimer and his co-authors write: "Setting aside or minimizing the importance of key structural uncertainties in underlying processes is a frequent outcome of the drive for consensus," and, therefore, "it is … important that policymakers understand the more extreme possibilities that consensus may exclude or downplay." [Oppenheimer et al. 2007] Indeed, it has been predicted that high values of ECS will be assigned higher probability in AR5 than in AR4, as more of the existing dissonance emerges. [Trenberth 2010]

One can anticipate that those responsible for contributing to AR5 will hold a wide range of credible views about the strengths of the feedbacks in the climate system and therefore

about the magnitude of the surface temperature change that will accompany a given concentration change. It would be unfortunate if this range of views were again collapsed into a single distribution of the ECS without extensive discussion. [InterAcademy Council 2010] Indeed, there are several ways for AR5 to go beyond AR4 in its reporting of disagreement. At the simplest level, the writers of AR5 could improve on AR4 by writing: "Some of us think the probability of the equilibrium climate sensitivity exceeding 4.5°C is 17% and some of us think it is 5%, and, in some detail, here's why." Users like me would learn that the authors of AR5, a group that arguably at the time of writing will know as much as any group in the world about ECS, could not converge on a single representation of the high end of its probability distribution. We would gain insight into the obstacles that prevented convergence within the group. We would develop a fuller understanding than was available in AR4 regarding what scientists know and do not know about high-consequence events.

The first stages of the process will probably not differ from those pursued in AR4. A group of experts will review the literature and will solicit the views of individuals. Each expert will do his or her best to create a personal synthesis and then, as part of the give and take of joint authorship, will undertake to persuade others to see the problem the same way. Disagreements will remain, and consensus will not emerge. The issue is what happens after that. The richness of climate science lies in the remaining disagreement. This disagreement must be fully reported, if users in the policy community are to be well served.

Early steps toward exploring the dissonance across analysts are available through a technique called expert elicitation. Pioneered in the environmental domain by Granger Morgan, expert elicitation creates formal settings where leaders in a field, privately and anonymously, provide quantitative answers to specific technical questions. [Morgan and Henrion 1990] The questions are often designed to produce quantitative estimates for a concept (like the probability of some event or the resultant level of damage) that experts usually discuss without quantification. Indeed, one recent exercise from the Morgan group specifically explores experts' estimates of the probability distribution for the equilibrium climate sensitivity. [Zickfeld et al. 2010] To be sure, the results of an expert elicitation are no substitute for the collective assessment process performed in the writing of an IPCC chapter, which features an exchange of views and (at least ideally) the full consideration of available information. But expert elicitation has a place in AR5, it seems to me, because its processes are internally disciplined and because its results can provide handholds for the authors of a chapter seeking to communicate the differences in the views that remain after the group's processes have run their course. The option of including expert elicitation has been argued for a long time. [Moss and Schneider 2000]

A less formal method of acquiring insight into differences across individual views can be developed using exploratory data analysis to study compilations of individual views such as the one presented in Fig. 2 above. Presumably, by including Fig. 2, the AR4 authors intended to communicate that a diversity of individual views about the ECS distribution co-existed at the time of writing. But although the text accompanying the figure in AR4 reports the underlying assumptions for each distribution, it does not present any quantitative analysis that looks *across* the distributions.

To illustrate what AR5 might do, I have played with a data set of eleven distribution functions for the equilibrium density function developed and displayed by Malte Meinshausen [Meinshausen 2006, Figure 28.1, p. 267]. Figure 3 presents these eleven distributions as cumulative distributions, which are the integrated versions of probability density functions. The set of eleven distributions is similar to the AR4 data set underlying Fig. 2; both were compiled at approximately the same time.
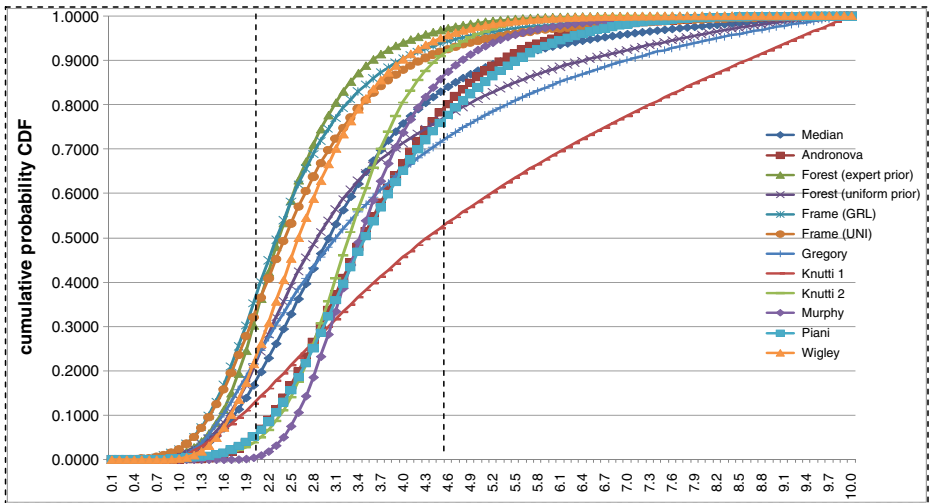
Fig. 3 Eleven pdf's for the Equilibrium Climate Sensitivity [Meinshausen 2006] are displayed as cumulative distribution functions. Meinshausen cuts off all distributions at 10°C. IIASA authors used this data base [O'Neill et al. 2010] and made it available as an Excel file at http://www.iiasa.ac.at/~riahi/Interim_Targets/; see the spreadsheet labeled "climate sensitivity PDFs." As in Fig. 2, vertical lines at 2°C and 4.5°C have been added to assist the reader

To shed light on how AR4 reported the probability that the equilibrium climate sensitivity exceeds 4.5°C, consider how this probability is viewed by Meinshausen's eleven analysts. An inspection of Fig. 3 shows a bimodal grouping. Frame assigns 72% probability to ECS being lower than 4.5°C, and thus 28% probability to ECS being higher than 4.5°C. Actually, four of the eleven analysts locate this latter probability between 22% and 28%, five locate it between 3% and 9%, one locates it in the middle at 14%, and an outlier locates it at 48%. That's a huge range of views. Is it any wonder that these AR4 authors found themselves tied in knots when they tried to report a collective view while not allowing themselves the option of disclosing their internal disagreements?

For the record, the spread across the eleven estimated probabilities that the ECS is *less than 2.0°C* is even more bimodal: four between 0% and 5%, six between 20% and 35%, only one in the middle at 13%. Actually, the second group of six can be described as two subgroups of three each, the first three with values between 20% and 22% and the second three with values between 30% and 35%. Moreover, the eleven estimates of the two probabilities (less than 2°C, more than 4.5°C) are weakly correlated: the correlation coefficient ("R-squared" value) for the eleven pairs of points is only 0.37.

This limited exploration, I hope, conveys some of the flavor of analyses that might be included in AR5 to convey the richness of the disagreements among analysts. AR5 would presumably not assign the eleven models equal weight, because the authors would know (as I do not) that some of the models are not actually independent. One can imagine that, even after correcting for lack of perfect independence, the apparently bimodal character of the responses in the Meinshausen data set will emerge once again in the literature that AR5 surveys. Suppose, furthermore, that AR5 then uses formal statistical tests and confirms that bimodality is a significant feature of the set of pdf's for ESC. AR5 would then be able to report that "two camps co-exist"—a stronger and more revealing statement about the state of the science than "consensus was difficult to achieve." And AR5 would explain *why* these

camps exist: whether the "priors" are different, whether paleoclimate constraints are used differently, whether particular feedbacks are suppressed or not. AR5 would also discuss how new research might test particular hypotheses critical to specific camps.

To be sure, one must guard against an excessive focus on the equilibrium climate sensitivity. One weakness of the ECS is that it assumes that as temperature reaches equilibrium the carbon cycle is unchanged. Thereby, ESC does not capture the full range of response to emissions. [IPCC 2007c, Section 3.5.3.1] For AR5 to do justice to high-consequence events, it needs somehow to include carbon-cycle feedbacks, which have high leverage on outcomes and have their own uncertainties. Martin Manning suggests that AR5 might contain a two-phase analysis. Phase 1 would deal with ECS, which by definition assumes that CO_2 is held fixed at twice its preindustrial concentration by an emissions trajectory that is tuned to compensate for sources and sinks. Phase 2 would explore carbon-cycle feedbacks that alter these sources and sinks, like forest degradation and outgassing from tundra, and thereby add complexity to the relationship between CO_2 concentration and surface temperature. [Martin Manning, private communication]

A second weakness of ECS is that it emerges from models only after the earth system has responded to emissions for centuries. [Solomon et al. 2010] Arguably, outcomes manifested during the next 70 years are more relevant to policy making than multi-century outcomes. Accordingly, some climate modelers prefer to discuss outcomes using the transient climate response (TCR) instead of ECS when communicating the risks of climate change in the IPCC context. [Meinshausen et al. 2009] TCR is tightly defined as "the annual global mean temperature change at the time of CO_2 doubling in a climate simulation with a 1% per year compounded increase in CO2 concentration." [IPCC, 2007b, p. 718] In AR4 three TCR probability distributions are shown, each based on an analysis of volcano data. The associated text states that 95% of the TCR distribution lies between 1.0°C and 3.5°C. [IPCC 2007b, Figure 9.21]

One can imagine a detailed discussion in AR5 of TCR, ECS on its own, and ECS with CO_2-cycle feedbacks. AR5 would explain the major reasons for the wide probability distributions associated with all three temperatures and would clarify how each of the three concepts bears on various policy time-horizons.

5 The AR5 guidance notes: comments and suggestions

The authors of the essays in this special issue of *Climatic Change* were asked to comment on a guidance document prepared for the authors of AR5. [Mastrandrea et al. 2010, henceforth "AR5 Guidance Notes" or, sometimes, "Notes"] The text of the Notes states: "These notes define a common approach and calibrated language that can be used broadly for developing expert judgments and for evaluating and communicating the degree of certainty in findings of the assessment process." I hope that revised AR5 Guidance Notes will take into account both of the main points in this essay: 1) AR5 should present vivid descriptions of high-consequence outcomes, and 2) AR5 should inform the reader about residual disagreements among authors. [3]

From the story of "cannot be excluded" told above (see Fig. 1 and surrounding text), I extract that it will also be important for the revised AR5 Guidance Notes to provide explicit

---

[3] I have learned that the Guidance Notes are hard to revise. But the whole point of this volume of essays is to improve guidance. Thus, when I refer to "revised" Notes, I include supplements and annexes and all other strategies that the IPCC could use to improve guidance.

guidance about *silence*. It is never easy to decide when to communicate and when to remain silent. In the absence of an external pressure to do otherwise, the record shows that groups of experts will choose silence when confronted with outcomes of low probability for which scientific understanding is preliminary, even when these outcomes have high consequences. Silence in these situations has serious negative consequences for public discourse. I recommend that the Notes strongly urge *vivid description* as the default value. The Notes should affirm that there are overriding obligations to the policy community, which wishes to learn what is important, whether poorly known or well known.

When knowledge is preliminary, it is also usually controversial, and the AR5 authors will need guidance regarding how to summarize discordant views. The AR5 Guidance Notes should help the authors distinguish synthesis from consensus. Disclosing only consensus should not be the objective. Rather, producing a synthesis, one that presents not only what everyone can agree upon but also important residual disagreements, is the objective. Right now, this important instruction is indeed written down, but is relatively well hidden away in the Tasks and Responsibilities Annex of the Appendix on Procedures of the document, "Principles Governing IPCC Work":

> The essence of the Lead Authors' task is synthesis of material drawn from available literature as defined in Section 4.2.3. ...Lead Authors are required to record in the Report views which cannot be reconciled with a consensus view but which are nonetheless scientifically or technically valid. [IPCC, 1999, Appendix A, Annex 1, p. 11]

Oppenheimer and his co-authors make a similar point: "Increased transparency, including a thorough narrative report on the range of views expressed by panel members, emphasizing areas of disagreement that arose during the assessment, would provide a more robust evaluation of risk." [Oppenheimer et al., 2007] Such transparency would increase the value of IPCC reports to the policy community.

Already, the concept labeled "agreement" plays a central role in the AR5 Guidance Notes. It is one of the two axes of its only figure; the other axis is "evidence." The objective of this figure is to demonstrate that "confidence" in a finding is bolstered by "consistency of evidence" and "degree of agreement." "Agreement," never actually defined in the Notes, is intended to capture the degree of consensus in the scientific community and the extent to which there are competing explanations. An earlier IPCC reports sows confusion by suggesting that "agreement" refer to "the extent to which the sources of information point in the same direction," a definition that better describes "consistency." [IPCC 2002, Section 2.3.1, Table 2.2 within Box 2.1, "Risk and uncertainty vocabulary used in this report."]

Some might find it troublesome that the two core concepts proposed in the Notes, "consistency" and "agreement," are strongly correlated, inasmuch as robust evidence and consensus tend to go hand in hand. Let's explore this further. Where might one find either consistent evidence and weak agreement or inconsistent evidence and strong agreement?

The combination of *robust* evidence and *low* agreement requires many explanations to be in play, even though there is a diversity of high-quality evidence. Such situations can arise when key information is missing. Even though much is known about the ice ages, many alternative feedback mechanisms are plausible. And even though much is known about sea level change, the lack of information about key ice sheet processes allows disparate projections of future sea level to co-exist. As for the combination of *limited* evidence and *high* agreement, it describes cases where experts have uniformly converged on a single explanation even though robust evidence is lacking. One would expect this combination to be rare: the more limited the data, usually, the larger the number of hypotheses in play. However, evidence that is limited but consistent with some disastrous

outcome could be expected to elicit wide agreement about the desirability of precautionary action. Clearly, not only "agreement" but also "consistency" could benefit from more precise discussion: there is a need to distinguish situations where consistent data are ample from situations where they are sparse.

In short, the presentation of "agreement" and "consistency" in the current AR5 Guidance Notes would benefit from supplementary material.

It will take courage to disclose lack of consensus, just as it will take courage to describe poorly understood high-consequence outcomes vividly. But the coexistence of contending views ("low agreement") is normal in science, not a cause for embarrassment, and users of the report need this information. Indeed, once IPCC authors accept that full reporting of competing views is a virtue, the evolution of understanding can be emphasized across successive Assessment Reports. Every chapter of AR5 can more fully comment on the views presented in the corresponding chapter of earlier reports and explain how recent research has altered perspectives. AR5 authors can also look forward, explaining how new research results could improve their understanding and speculating about what will be known in 5, 10 and 20 years. In subsequent reports, these predictions can be compared with the new state of knowledge. Among the beneficiaries of such discussion will be policy makers, who, if they are to implement "iterative risk management," need not only insight into what is known and unknown, but also insight into the rate of learning. (National Research Council, 2011) The AR5 Guidance Notes should do all they can to generate a report that presents earth systems science as an evolving human enterprise.

In the Appendix, I suggest additional revisions of the AR5 Guidance Notes that are more technical.

6 What will be known about high-consequence outcomes, long after AR5?

This article urges that the IPCC give greater priority to high-consequence outcomes. Looking beyond AR5, such a prioritization encompasses not only what is written in IPCC reports but also what is studied by the earth science community. No matter how superb the communication, it cannot compensate for poor understanding. Because high-consequence outcomes matter greatly to policymakers, one must consider whether the fraction of R&D dedicated to high-consequence outcomes is sufficiently high.

In AR5, the IPCC must help the policymaker understand the current incompleteness of earth systems science. The policymaker needs to learn about the full range of outcomes, with probabilities attached and disagreements among experts revealed. Some policymakers will decide that even if high-consequence outcomes could be ruled out, bold steps are warranted. Others may be motivated largely by the desirability of avoiding disasters, no matter how low their probability. AR5 can set a standard for full disclosure that subsequent Assessment Reports will emulate.

Appendix: five recommendations bearing on Table 1 of the AR5 Guidance Notes

In this appendix I make five recommendations regarding revisions of Table 1 of the AR5 Guidance Notes, reproduced here as Table 1. Some readers may consider the recommendations unnecessary, on the grounds that each of the recommended practices is already universally understood and followed. My direct experience, however, contradicts this.

All five recommendations are technical points about statements of the form "$x > A$ is likely" and "$B > x > A$ is likely" I argue for precision in defining "likely," and its partners. Achieving precision in this narrow sense removes unnecessary ambiguity and fosters coherent discussion by focusing all discussion of uncertainty on physical values (here, temperatures) at boundaries like A and B.

First, a technical point about "likely" versus "very likely" and similar distinctions. In Table 1, "likely" is associated with 66% to 100% probability and "very likely" with 90% to 100% probability. A critical piece of further guidance is needed. Suppose the question is the anticipated height at maturity of a child of six: a girl has a certain height and the question is whether she will end up taller than five foot one. Suppose one pediatrician would assign 93% probability to this outcome. You could say either that the pediatrician thought it was "likely" that the girl would become taller than five foot one, or that the pediatrician thought it was "very likely." Both are correct statements. But you must report this judgment as "very likely," if the hard-earned distinctions in the Guidance Notes are to have any meaning. Otherwise, disagreement can masquerade as consensus. Suppose another pediatrician would assign 81% probability to the same outcome (becoming taller than five foot one). The proper summary is not to write that both believe that the outcome is "likely," but rather to write: "One would assign "likely" and other "very likely" to this outcome, and here's why."

The second technical point concerns two contexts for the statement, "$x > A$." This statement can arrive on your doorstep, with A not at your disposal; the statement could be that a specific six-year-old girl will be at least five foot five. The only task available to you is to judge its probability of being true. You can't change A. Let's say you decide that the probability that the girl will be at least five foot five is 75%. You then inspect Table 1, below, and learn that such a statement needs to be reported as "likely." End of the matter.

More often, however, the context for "$x > A$" leaves you free to choose A. Table 1 disciplines your choice. You must choose an A associated with one of the boundary

Table 1 Likelihood scale proposed in the AR5 Guidance Notes [Mastrandrea et al. 2010, p. 5]	Term	Likelihood of the outcome
	Virtually certain	99–100% probability
	Very likely	90–100% probability
	Likely	66–100% probability
	About as likely as not	33 to 66% probability
	Unlikely	0–33% probability
	Very unlikely	0–10% probability
	Exceptionally unlikely	0-1% probability

probabilities, 90%, for example (the lower bound of "very likely"). You should not report that "*the girl is very likely to reach five foot one,*" if you consider that outcome to have 93% probability. Rather, you must report the height (greater than five foot one, say five foot three) that the girl has 90% probability of exceeding at maturity. The reader has a right to infer that you are assigning a probability of 90% to any statement being "very likely," not 93%. The reader has a right to assume that you have embarked on an intelligent exercise, not a capricious one.

The third technical point is analogous to the second point, but for two-sided ranges, "$\underline{B > x > A}$," that contain some but not all of the possible values of *x*. If the range arrives at your doorstep, you can decide only its qualitative descriptor ("likely" vs. "very likely," for example), but you cannot shift *B* or *A*. However, if you are in charge of choosing *B* and *A*, and, for example, you are seeking values that will provide a "likely" range, you must choose a *B* and an *A* that include 66% but not, say, 75% of the expected values of *x*. Your reader will assume that you are being as precise as you can, not that you prefer to be vague.

In this case, let's dispense with our analogy to the growing girl, because this point is particularly germane to AR5. If AR5 reports that the equilibrium climate sensitivity, for example, "is likely to be between 1.8°C and 5.0°C," the meaning of 'likely" must correspond to 66%, Yet, I have already heard people with responsibility for AR5 declare that "likely" in such statements will correspond to *any* percent between 66% and 90%. I fervently hope that the AR5 authors can be self-conscious enough to avoid such completely unnecessary fuzziness. Choose the boundaries of the "likely" range corresponding to 66%, making your best effort, and then discuss in the text alternative choices for the boundaries.

The fourth technical point is a further constraint on two-sided ranges, "$\underline{B > x > A.}$" You must center all such ranges, i.e., you must choose an *A* and a *B* such that the probabilities of *x* falling below *A* and above *B* are equal, since this always an available option. If you report that the girl has a likely height (66% chance) at maturity between five foot six and five feet nine, the reader has a right to assume that you have taken the trouble to choose these boundaries such that the chances of her someday being less than five foot six and more than five foot nine are equal. Unless otherwise reported, all two-sided ranges reported in IPCC documents must be centered.

The whole likely/very-likely apparatus must not become a straightjacket. One can imagine a situation where a group wishes to report asymmetric boundaries for some range—assigning, for example, 17% to the probability of a value falling below the lower end of the range and 5% to the probability of its falling above its upper end (thus, an *uncentered* range containing 78% of the possible values); indeed, footnote 2 identifies an instance where, perhaps, some experts wished to report a range of this sort. Rules must not exclude the reporting of an asymmetric range, fully explained.

The fifth technical point is the observation that the probability of the statement "$x > A$" is not the same for the two statements: "$x > A$ is very likely" and "$B > x > A$ is very likely." The two statements, respectively, describe one-sided and two-sided ranges. When "$x > A$" is "very likely," there is a 10% chance that x is less than *A*, because there is a 90% chance that *x* is greater than *A*. When "$B > x > A$" is "very likely," there is a 5% chance that *x* is less than *A* (and a 5% chance that *x* is greater than *B*). The girl is very likely to become (has a 90% chance of being) taller than five foot three; her future height is very likely to fall (has a 90% chance of falling) between four foot eleven and six foot one (range boundaries defining 5% and 95%).

In my opinion, all five technical points need explication in a revised AR5 Guidance Notes. IPCC writers should use the most restrictive word—always "very likely," for example, when a statement is both "likely" and "very likely." They should choose the highest possible correct value for A when asserting that $x > A$ is likely. They should center all two-sided ranges. And they should clearly distinguish one-sided and two-sided inequalities. Whenever there is a compelling reason not to follow one of these practices, the reader should be informed and told why.

References

InterAcademy Council (2010) Climate change assessments, review of the processes and procedures of the IPCC. available at: http://reviewipcc.interacademycouncil.net/

IPCC (1999) Principles Governing IPCC Work Procedures for the Preparation, Review, Acceptance, Adoption, Approval and Publication of IPCC Reports. Adopted at the Fifteenth Session (San Jose, 15–18 April 1999) and amended at the Twentieth Session (Paris, 19–21 February 2003) and Twenty-first Session (Vienna, 3 and 6-7 November 2003)

IPCC (2002) Climate Change 2002, Mitigation of Climate Change. Contribution of Working Group III to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, New York

IPCC (2007a) *Climate change 2007: Synthesis Report*. Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, New York

IPCC (2007b) 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, New York.

IPCC (2007c) Climate change 2007: mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, New York

Mastrandrea MD, Field CB, Stocker TF, Edenhofer O, Ebi KL, Frame DJ, Held H, Kriegler E, Mach KJ, Matschoss PR, Plattner G-K, Yohe GW, Zwiers FW (2010) Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *Intergovernmental Panel on Climate Change (IPCC).* Available at http://www.ipcc-wg2.gov/meetings/CGCs/Uncertainties-GN_IPCCbrochure_lo.pdf. Also referred to in the text as "AR5 Guidance Notes" and "Notes."

Meinshausen M (2006) What does a 2°C target mean for greenhouse gas concentrations? A brief analysis based on multi-gas emission pathways and several climate sensitivity uncertainty estimates. In: Schellnhuber JS, Cramer W, Nakicenovic N, Wigley TML, Yohe G (eds) Avoiding dangerous climate change. Cambridge University Press, Cambridge

Meinshausen M, Meinshausen N, Hare W, Raper SCB, Frieler K, Knutti R, Frame DJ, Allen MR (2009) Greenhouse-gas emission targets for limiting global warming to 2°C. Nature 458:1158–1163

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

Morgan MG, Keith D (2008) Improving the way we think about projecting future energy use and emissions of carbon dioxide. Climatic Change 90:189–215

Moss R, Schneider S (2000) Uncertainties. In: Pachauri R, Taniguchi T, Tanaka K (eds) Guidance papers on the cross cutting issues of the third assessment report of the IPCC. Intergovernmental Panel on Climate Change (IPCC), Geneva

National Research Council (2011) *America's climate choices*. Washington DC: National Academies Press. Four related panel reports have also appeared. See http://americasclimatechoices.org/

O'Neill BC et al (2010) Mitigation implications of midcentury targets that preserve long-term climate policy options. PNAS 107(3):1011–1016

Oppenheimer M et al. (2007) The limits of consensus. Science, vol 317, September 14, 2007, pp. 1505-6. See also: Solomon, S., et al, 2008. "A closer look at the IPCC Report." Science 319: 409-410, 25 January 2008, and Oppenheimer, M. et al (2008). Response to "A closer look at the IPCC Report." Science 319: 410, 25 January 2008

Socolow R (1976) Failures of discourse: obstacles to the integration of environmental values into natural resources policy. In: Tribe LH, Schelling CS, Voss J (eds) When values conflict: essays on environmental analysis, discourse, and decision. Ballinger, Cambridge

Solomon S et al (2010) Climate stabilization targets: emissions, concentrations, and impacts over decades to millennia. National Academies Press, Washington

Trenberth K (2010) More knowledge, less certainty. Nature Reports Climate Change. http://www.nature.com/climate/2010/1002/full/climate.2010.06.html

Weitzman M (2009) On modeling and interpreting the economics of catastrophic climate change. Review of Economics and Statistics 91(1):1–19, (Feb 2009). See also "Revisiting fat-tailed uncertainty in the economics of climate change," *Review of Environmental Economics and Policy*, forthcoming

Zickfeld K et al. (2010) Expert judgments about transient climate response to alternative future trajectories of radiative forcing. Proceedings of the National Academy of Sciences 107: 12451-12456, www.pnas.org/cgi/doi/10.1073/pnas.0908906107.