Conceptualizing Uncertainty: An Assessment of the Uncertainty Framework of the Intergovernmental Panel on Climate Change

Forthcoming in: European Studies in Philosophy of Science. Recent Developments in the Philosophy of Science – EPSA15 Düsseldorf

Nicolas Wüthrich
Department for Philosophy, Logic, and Scientific Method
London School of Economics and Political Science
N.Wuethrich@lse.ac.uk

Abstract

We are facing uncertainties regarding climate change and its impacts. To conceptualize and communicate these uncertainties to policy makers, the Intergovernmental Panel on Climate Change (IPCC) has introduced an uncertainty framework. In this paper, I assess the latest, most developed, version of this framework. First, I provide an interpretation of this framework, which draws from supporting documents and the practice of its users. Second, I argue that even a charitable interpretation exhibits three substantial conceptual problems. These problems point towards untenable assumptions regarding evidence aggregation in the context of climate scientific findings. Third, I put forward a tentative roadmap for improving the uncertainty framework.

Key words: Climate science, Evidence aggregation, IPCC assessment report, Uncertainty guidance note

1. Introduction

At the Paris Climate Change Conference in November 2015, an agreement was reached which is regarded as a milestone in addressing global climate change. The parties committed to limit the increase in the global average temperature to below two degrees Celsius above pre-industrial level (United Nations 2015, Article 2). It appears that the delegates acted upon President Holland's address in which he claimed that the hope of all of humanities rests on the shoulders of the conference participants (Hollande 2015, 6). In light of these high stakes, it can be hoped that the agreement is based on the best available scientific knowledge about the causes and consequences of climate change. Indeed, in the Paris Agreement an effective and progressive response to climate change is explicitly asked for "in light of the best available scientific knowledge" (United Nations 2015, Preamble).

The Intergovernmental Panel on Climate Change (IPCC) synthesizes the current state of knowledge about climate change (Mastrandrea et al. 2011, 676). This involves understanding and being able to communicate the uncertainties surrounding these scientific findings. To achieve these ends, the IPCC developed an uncertainty framework (ibid.).

The uncertainty framework is best understood as an ambitious attempt at a unified conceptualization of different types of uncertainties, including model and data uncertainty, scenario uncertainty, as well as ethical uncertainty (IPCC 2013, 138; Mastrandrea et al. 2011, 676). The framework is in constant development with later versions building on previous ones. The latest, and most developed, version of the framework equips scientists with a confidence and likelihood metric to qualify their statements.

There exists already a literature addressing previous and current versions of the IPCC's uncertainty framework (see Adler and Hirsch Hadorn 2014 for a review).

The majority of the literature focuses on issues surrounding the interpretation of probabilistic information by the readers of the IPCC reports (e.g. Budescu et al. 2014; Morgan 2014). However, the engagement with the conceptual foundations of the latest version of the uncertainty framework remains limited (see Aven and Renn 2015; Jones 2011; Socolow 2011). In this paper, I attempt to close this gap.

The paper is structured as follows. First, I provide an interpretation of the framework which draws from introductory documents and the actual practice of its users (section 2). With this interpretation in hand, I, second, identify three conceptual problems which point towards untenable assumptions regarding evidence aggregation in the context of climate scientific findings (section 3). Third, I put forward a tentative roadmap for improving the uncertainty framework (section 4).

2. The Uncertainty Framework of the IPCC: An Interpretation

The IPCC was established by the United Nations Environment Programme and the World Meteorological Organization in 1988 (IPCC 2014a, 1). The IPCC is not conducting original research but synthesises the current scientific knowledge about climate change. This synthesis is presented in assessment reports (ibid.). Each assessment report has three parts, written by one working group respectively. Part I covers the physical aspect of the climate system and its change (IPCC 2014b, 1). Part II assesses the vulnerability of socio-economic and natural system to climate change and processes of adaption (ibid.). Part III reviews options for mitigating climate change (ibid.).

To adequately account for the uncertainties involved in scientific findings, the IPCC uses in its latest assessment report an updated version of the uncertainty framework. This framework serves two functions. It is an analytical instrument to understand uncertainty and a tool for communicating uncertainties to policy makers (Mastrandrea et al. 2010, 1).

Throughout this paper, I will focus on the assessment report of working group I. The IPCC provides two supporting documents which explain the framework: the "Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties" (Mastrandrea et al. 2010) and a commentary article by Mastrandrea et al. (2011).[1] I also take into account the practice of the authors of the assessment reports. As the discussion below will make clear, the supporting documents and the practice reveal ambiguities and inconsistencies in the framework. Hence, the aim of this section consists in providing a charitable interpretation of the framework, which dissolves the ambiguities in a way that minimizes inconsistencies.

To get a handle on the uncertainty framework, consider the following examples of its application. There are cases in which solely a confidence term is used to characterize a finding:

> "The release of CO2 or CH4 to the atmosphere from thawing permafrost carbon stocks over the 21st century is assessed to be in the range of 50 to 250 GtC [giga tons of carbon] (…) (*low confidence*)." (IPCC 2013, 27, my emphasis)

There are cases in which solely a likelihood term is used:

> "It is *likely* that the frequency of heat waves has increased in large parts of Europe, Asia and Australia [since the 1950s]." (ibid., 5, my emphasis)

Finally, there are cases in which both confidence and likelihood terms are used:

> "In the Northern Hemisphere, 1983-2012 was *likely* the warmest 30-year period of the last 1400 years (*medium confidence*)." (ibid., 3, my emphasis)

---

[1] A comparison of the current version of the uncertainty framework, published in 2010 and used in the Fifth Assessment Report, with the previous one, published in 2005 (and used in the Fourth Assessment Report published 2007), can be found in Mastrandrea et al. 2010, Annex A.

How should these confidence and likelihood qualifications be understood and how are they related to each other?[2]

2.1 Likelihood terms

Let me begin with the likelihood terms. The guidance note reveals that these terms correspond to ranges of probabilities: virtually certain (probability of the occurrence of the outcome is 99-100%), very likely (90-100%), likely (66-100%), about as likely as not (33-66%), unlikely (0-33%), very unlikely (0-10%), and exceptionally unlikely (0-1%) (Mastrandrea et al. 2010, 3). Accordingly, these likelihood terms express a quantitative measure of uncertainty (ibid., 1).

The basis for the ascription of probabilistic information can be statistical or modelling analysis or elicitation of expert views (ibid., 3). Model analysis involves the generation and analysis of time series data for a variable (e.g. global mean surface temperature) and a period of interest (e.g. 1980-2050) from single or multiple models. Expert elicitation techniques are procedures which aim at determining experts' subjective degree of belief for the values of interest. They are primarily used to capture meta-knowledge of experts about limitations of climate models and observational data (Morgan 2014, 7176).

2.2 Confidence terms

Let me now turn to the confidence terms. In contrast to probabilistic information, confidence is expressed qualitatively, i.e. it can be very high, high, medium, low, or very low (Mastrandrea et al. 2010, 1). The guidance note gives the following indication for arriving at these confidence statements (ibid.):

[2] There are also cases in which no qualifying term is used. The absence of qualifying terms indicates a finding for which "evidence and understanding are overwhelming" (Mastrandrea et al. 2010, 2).

> "Confidence in the validity of a finding. Based on the type, amount, quality, and consistency of *evidence* (e.g. mechanistic understanding, theory, data, models, expert judgment) and the degree of *agreement*. Confidence is expressed qualitatively." (ibid., my emphasis)

This statement needs to be disentangled. I begin with the two notions of 'evidence' and 'agreement'. I then shed light on how these two notions are combined to arrive at confidence statements.

The supporting documents suggest assessing the available evidence on the basis of the type, amount, quality, and consistency of evidence. These four dimensions are best viewed as criteria which should be considered individually by the authors (ibid., 2).

Under the heading of *type of evidence*, the guidance note provides five categories of evidence: mechanistic understanding, theory, data, models, and expert judgement (ibid., 1). Mechanistic understanding is described as understanding of the physical processes governing a particular phenomenon (Mastrandrea et al. 2011, 678). As Jones (2011, 737) notes, it is puzzling why theory is treated as one category of evidence, since, normally, evidence is viewed to be (dis)confirming theories. A charitable way of understanding theory as a subcategory of evidence is to see it as theoretical knowledge supporting (or undermining) particular explanations or predictions which are reported in a finding. Predictions can be supported by theory if the predictions are model-based (and the theory supports the structural assumptions of the model) or predictions are based on expert elicitation, where experts ground their judgements in theory. This interpretation of theory as an evidence category is suggested by the practice of the authors:

> "In summary, *ice-dynamics theory*, numerical simulation, and paleo records indicate that the existence of a marine-ice sheet instability (…) is possible in response to climate forcing." (IPCC 2013, 1174, my emphasis)

The terms 'amount' and 'quality of evidence' are not defined in the supporting documents. Authors are using *amount of evidence* to denote different things: a) number of observation points (e.g. IPCC 2013, 137, 158); b) number of models or total amount of scenarios run on selected models (e.g. simulation of Greenland ice sheet, Figure 5.16, ibid., 428); or c) the number of studies (e.g. ibid., 129). Q*uality of evidence* is used by the authors in relation to observational data and models. Data quality involves judgements about instrument design, equipment handling, or data processing (e.g. ibid., 143). Model quality is assessed based on empirical model performance and adequate representation of relevant causal factors (ibid., 749, 753f.).

Consistency of evidence is defined as "(…) the extent to which it [i.e. evidence] supports single or competing explanations of the same phenomena, or the extent to which projected future outcomes are similar or divergent" (Mastrandrea et al. 2011, 678).

Evidence is expressed on a qualitative scale: evidence can be robust, medium, or limited (Mastrandrea et al. 2010, 2). *Robust evidence* is defined as multiple, consistent independent lines of high-quality evidence (Mastrandrea et al. 2011, 678). Notice that here an additional criterion for the evaluation enters the scene which is not explicitly introduced in the framework: the (in)dependence of different pieces of evidence. For example, the models in an ensemble can be independent to a higher or lower degree given how many model components they share (IPCC 2013, 755). The supporting documents do not define the levels of *medium* and *limited evidence*. In particular, there are no aggregation rules given which might indicate the relative importance of type, consistency, independence, amount, and quality of evidence. The practice of the authors does not reveal specific aggregation rules or principles either.

Let me now turn to the notion of *agreement*. Agreement is expressed qualitatively: agreement can be high, medium, or low (ibid., 3). Agreement is not defined in the guidance note. However, Mastrandrea et al. (2011) offer the following two accounts of agreement in their commentary:

> "[Agreement] is the level of consensus in the scientific community on a particular finding." (ibid., 678).

> "[Agreement indicates] the degree to which a finding follows from established, competing, or speculative scientific explanations." (ibid.)

At first sight, these two definitions differ. A salient way of reconciling them would be to add the assumption that the level of consensus in a scientific community depends on the degree to which a finding follows from established, competing, or speculative scientific explanations. Given this assumption, the second definition entails the first one.

Interpreting in the way suggested here leads to a problem: agreement and consistency cannot be ascribed independently from each other. Recall that both agreement and consistency are defined with respect to a finding in the assessment report. However, if agreement is low (due to the presence of competing explanations), then consistency needs to be low as well, since consistency is defined as the amount of explanations supported by the evidence. The same holds for high agreement and high consistency. Mastrandrea et al. (2011) seem to sense this tension and explain the difference between agreement and consistency as follows:

> "Agreement is not equivalent to consistency. Whether or not consistent evidence corresponds to a high degree of agreement is determined by other aspects of evidence such as its amount and quality; evidence can be consistent yet low in quality." (ibid., 678).

This explanation is not satisfactory since it is in tension with our first attempt of making sense of agreement. Agreement has been so far understood as depending only on the amount of competing explanations for a finding. Here, Mastrandrea et al. suggest that it depends also on the amount and quality of evidence. The best way to circumvent this problem is to view ‘agreement’ as an umbrella notion which covers two different concepts: agreement a*s degree of consensus in the scientific community* and agreement *as consistency of evidence*.

This interpretational move is supported by the practice of the authors. When the authors are making use of the uncertainty framework, then they interpret ‘agreement’ in the majority of cases as consistency of evidence. Let me give an example:

> “*High agreement among analyses* provides medium confidence that oxygen concentrations have decreased in the open ocean thermocline in many ocean regions since the 1960s.” (IPCC 2013, 52, my emphasis)

However, there are also instances where agreement is viewed as consensus in the scientific community:

> “Many semi-empirical model projections of global mean sea level rise are higher than process-based model projections (…), but there is *no consensus in the scientific community* about their reliability and there is thus low confidence in their projections.” (ibid., 26)

Let me take stock at this point. The hierarchy of the notions which are introduced by the uncertainty framework, given the interpretational ambiguities, can be visualised as follows (see figure 1):

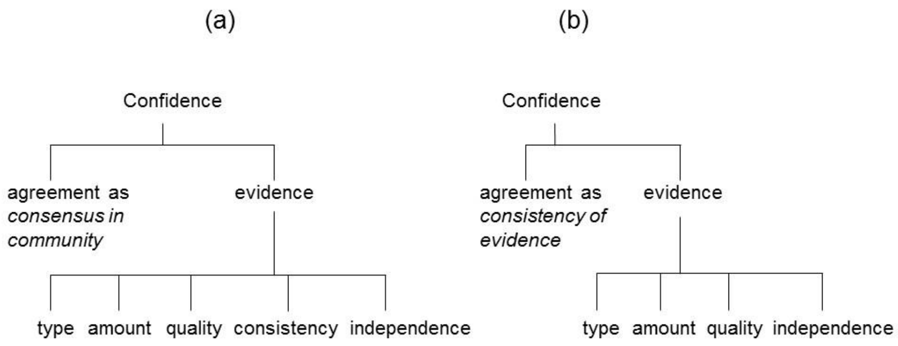

Figure 1: Elements of the confidence metric under two interpretations of agreement: (a) agreement as consensus in the scientific community, (b) agreement as consistency of evidence (my diagram).

After having discussed the evidence and agreement notions, let me now turn to their *aggregation into overall confidence statements*. The supporting documents specify that the increase in levels of agreement or evidence (individually, while holding the other level constant, or together) leads to a rise in the confidence level (Mastrandrea et al. 2010, 3). For findings with high agreement and robust evidence, the confidence level 'very high' should be assigned (ibid., 2). For findings with either high agreement or robust evidence a confidence level should be given if possible (i.e. high confidence or medium confidence). If this is not possible, then the summary terms should be used (e.g. robust evidence, medium agreement) (ibid., 3). For findings with low agreement and limited evidence, the summary terms should be used. Figure 2 visualizes these rules:

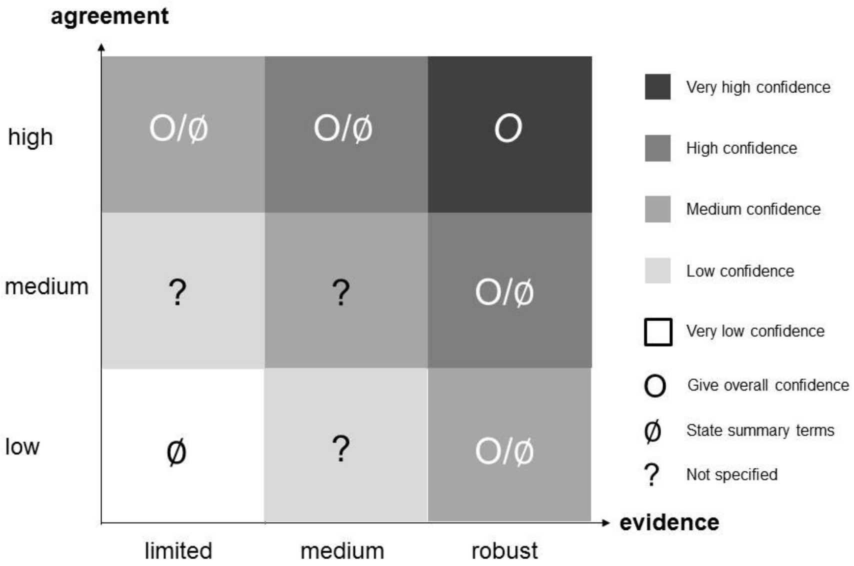

Figure 2: Aggregation of evidence and agreement into overall confidence statements (my diagram, based on Mastrandrea et al. 2010, 2-3).

So far, I have discussed the likelihood and confidence terms of the uncertainty framework. A crucial question remains: How are these two scales related to each other?

2.3 The relation between likelihood and confidence terms

Do likelihood and confidence terms convey the same or different types of information? The supporting documents do not rule out either of the two possible but incompatible answers.

To start, the likelihood metric can be interpreted as a quantified uncertainty tool which co-varies with the confidence metric. This interpretation would treat the likelihood and confidence metrics as substitutes, conveying the same information. The difference in the application of the two metrics would consist in the fact that not all types of evidence allow a quantified treatment of uncertainty. I refer to this reading as the *substitutional-interpretation*. The following statements from the supporting documents back this reading:

> "Depending on the nature of the evidence evaluated, teams have the option to quantify the uncertainty in the finding probabilistically." (Mastrandrea et al. 2010, 1)

> "[If] a range can be given for a variable, based on quantitative analysis or expert judgment: Assign likelihood or probability for that range when possible; *otherwise* only assign confidence" (ibid., 4, my emphasis)

In contrast, the confidence statements can be interpreted as meta-judgements and the likelihood statements as intra-finding judgements: The confidence metric allows assessing the goodness of the evidential basis of a finding whereas the likelihood metric can be used to specify the events which are mentioned in that finding. Under this reading, the two metrics would contain different information. This would allow the absence of co-variation between the likelihood and the confidence metric. I refer to this reading as the *non-substitutional interpretation*. The following statements from the supporting documents back this interpretation:

> "Author teams are instructed to make this evaluation of evidence and agreement the basis for any key finding, *even those that employ other calibrated language* (…)." (ibid., Annex A, my emphasis)

> "This scale [confidence metric] may be *supplemented* by a quantitative probability scale (…)." (ibid., Annex B, my emphasis)

How are the authors of the assessment reports dealing with this interpretational ambiguity? The practice reveals that in the clear majority of cases authors are opting for the non-substitutional interpretation. The following example illustrates this:

> "Estimates of the Equilibrium Climate Sensitivity (ECS) based on multiple and partly independent lines of evidence (…) indicate that there is *high confidence* that ECS is *extremely unlikely* to be less than 1°C (…)." (IPCC 2013, 871, my emphasis)

Finally, the supporting documents prescribe to use probabilistic information only if the confidence in the finding is high or very high (Mastrandrea et al. 2010, 4). No rationale is given for this rule. The implicit motivation could be that it is more problematic to assign probabilities given one has low confidence in the evidential basis for a finding. The authors of the assessment report seem to disregard this rule about the use of probabilistic information. There are multiple instances in which likelihood terms are used given very low, low, or medium confidence. Here is an example:

> "A nearly ice-free Arctic Ocean (…) in September before mid-century is *likely* under [emission scenario] RCP 8.5 (*medium confidence*)." (IPCC 2013, 92, my emphasis)

The salient way of reconciling the practice of the authors with the guidance note would be to interpret the rule as stating that *precise* probabilistic information, i.e. complete probability density functions, should only be given if confidence is high or very high.

This concludes my attempt at giving a coherent interpretation of the uncertainty framework. In the next section, the offered interpretation will serve as a background for an engagement with the conceptual foundations of the framework.

3. Conceptual Problems in the Uncertainty Framework

In this section, I argue that the uncertainty framework exhibits three substantial conceptual problems. These problems make clear that the current version of the framework is neither an adequate tool for conceptualizing the uncertainties involved in climate scientific findings nor for communicating them to policy makers.

The first problem concerns the *bifurcation of evidence and agreement in the confidence metric*. Given the two possible interpretations of agreement, this bifurcation does not uphold scrutiny. If one understands agreement as consensus in the scientific community, then the social fact of consensus should be a result of the evidence and should not be treated as an independent dimension. As thermometer readings should track temperature, the social fact of consensus in a community should supervene on the available evidence facts. Ultimately, it is the evidence which should guide our uncertainty assessment. If one understands agreement as consistency of evidence, then agreement is straightforwardly part of the evidence dimension and nothing separate from it.

This bifurcation between evidence and agreement leads to a second problem unfolding into a set of issues related to the *rules for aggregating evidence and agreement statements*.

To start, as illustrated in figure 2, the uncertainty framework allows for the combination of robust evidence and low or medium agreement. Recall that 'robust evidence' is defined as multiple, consistent independent lines of high-quality evidence (Mastrandrea et al. 2011, 678). If one understands agreement as consistency, then there cannot be low (or medium) agreement in light of robust evidence. Since robust evidence involves evidence that is consistent, agreement as consistency needs to be high in light of robust evidence. If one understands agreement as consensus in the scientific community, then it is puzzling how there can be a limited level of consensus in light of robust evidence given one makes the minimal assumption that scientists are basing their judgements on the available evidence.[3]

Even if one disregards this issue, a second issue emerges in relation to aggregation rules. The pairs 'limited evidence/high agreement' and 'robust evidence/low agreement' are treated symmetrically by getting assigned medium

[3] From a subjective Bayesian perspective one might argue that the combination of robust evidence and low agreement can be explained via different priors of experts. This strategy does not seem promising. The uncertainty framework *defines* robust evidence as "multiple, consistent independent lines of high-quality evidence" (Mastrandrea et al. 2011, 678). So verdict that we have *robust* evidence implies consensus among the scientists on quality as well as explanatory and predictive import of the evidence. This is incompatible with substantially different posteriors of experts, not matter what their priors were.

confidence. This symmetry is puzzling. The intuition that one faces less uncertainty given ‘robust evidence/low agreement’ than ‘limited evidence/high agreement’ seems natural. This intuition can be substantiated as follows. If one understands agreement as consensus in the scientific community, it seems questionable to give the fact of consensus the same weight as the evidential considerations. If one understands agreement as consistency, it seems problematic to give consistency considerations the same weight as the combined considerations about type, amount, quality, and independence of evidence. This point gains traction in the practice of the authors of the assessment report. Instead of weighting consistency in light of other evidential considerations, the authors *solely* use consistency considerations to arrive at overall confidence statements. Here is an example:

> “*High agreement among analyses provides medium confidence* that oxygen concentrations have decreased in the open ocean thermocline in many ocean regions since the 1960s.” (IPCC 2013, 52, my emphasis)

This constitutes bad epistemic practice since the other dimensions, which should enter a critical assessment of the underlying evidential basis for a finding, i.e. the type, amount, quality, and independence of the evidence, are not considered. If one assumes that the authors are considering these alternative dimensions implicitly, this practice is additionally not transparent to the readers of the report.

A final issue concerns the amount of information that is provided about the aggregation rules. Consider again figure 2. Why is it the case that sometimes one can give an overall confidence judgment and sometimes one is only supposed to give the summary terms? It is conceptually unclear how a line between these two classes of cases can be drawn. Furthermore, the diagonal in figure 2 is puzzling. The diagonal contains all matching evidence and agreement pairs (e.g. limited evidence, low agreement). In these clear cut cases, an overall confidence statement should be possible. However, as figure 2 illustrates, this is not the prescription of the framework.

The third problem can be located in the *rules about when it is permissible to use probabilistic information*. A charitable interpretation of the uncertainty framework states that only if the confidence is high or very high, precise probabilistic information should be used to express uncertainty. If one adopts the non-substitutional reading as suggested by the practice of the authors, it is not clear why one cannot use precise probabilistic information if the confidence in the evidential basis yielding this probabilistic information is low. To start, using probabilities in itself does not confer any epistemic merit on a finding. Furthermore, to prohibit communicating probabilistic information in these cases amounts to deliberately setting aside available information. This violates Carnap's plausible principle of total evidence (Carnap 1947), and, hence, should alert suspicion.

4. Concluding Remarks: A Roadmap for Improving the Uncertainty Framework

In section 2, I offered an interpretation of the uncertainty framework which drew from the supporting documents and the practice of the authors. The discussion made clear that the framework contains multiple ambiguities as well as inconsistencies. In section 3, I showed that even a charitable interpretation of the framework faces three substantial conceptual problems. In my point of view, these conceptual problems are sufficient to motivate a re-design of the framework.

The three conceptual problems point towards the issue of *evidence aggregation for climate scientific findings*. The IPCC assessment report of working group I involves paradigmatic cases for multi-modal evidence aggregation problems: Almost for every finding different lines of evidence (such as observations, model results, theory, and expert opinion) have to be combined. Accordingly, an adequate uncertainty framework for conceptualizing and communicating climate scientific findings should be based on an adequate evidence aggregation model for climate scientific findings. In light of the complexity of this aggregation task, I only sketch a roadmap for re-designing the uncertainty framework towards this goal. The roadmap contains three steps.

First, *the bifurcation between evidence and agreement in the confidence metric needs to be removed*. As my discussion has illustrated, agreement should not be viewed as an independent dimension from evidence. Rather, our confidence in a finding should be solely determined by the available evidence; the better the available evidence for a finding, the higher our confidence should be in this finding, and vice versa. Making this statement precise is the key task. Importantly, this is not to say that expert judgement plays no role in the process of evidence aggregation. However, the role of expert judgment, and the significance of expert agreement, should be spelled out in relation to the assessment criteria of evidence. The social epistemology literature on peer (dis-)agreement can be used as a starting point to clarify this further (e.g. Kelly 2005).

Second, *assessment criteria for the available evidence need to be identified and spelled out in sufficient detail*. The following questions should be answered: What criteria are relevant to assess the ensemble of available evidence? What criteria are relevant for assessing individual pieces of evidence? Can we define the assessment criteria in a formally precise way? The literature on evidence aggregation in medicine has already provided some conceptual resources along these lines (e.g. Clarke et al. 2014; Stegenga 2011).

Third, once the assessment criteria for evidence have been identified the *task of aggregating these criteria into overall confidence judgments* can be addressed. Here, the rich literature of social choice theory as well as multi-criteria decision analysis suggests itself as a source for technical tools. It remains to be explored to what degree general, non-case specific, aggregation rules can be developed for the context of climate scientific findings.

Let me close by noting that my discussion of the framework as well as the suggested roadmap has been solely taking into account uncertainties involved in the physical science basis of climate change. In particular, ethical uncertainty, which is addressed in working group III, has been deliberately set aside. In my point of view, the uncertainty framework has to be first improved to adequately address the uncertainties involved in the physical science basis of climate change before it can be generalized to different types of uncertainties.

Acknowledgments

Richard Bradley, Thomas Clarmann, Roman Frigg, Stephan Güttinger, Casey Helgeson, Gertrude Hirsch Hadorn, Vivek Sant, and Philippe van Basshuysen provided helpful feedback on earlier versions of this paper. I am also thankful for comments by two anonymous reviewers and audiences at the EPSA conference in Düsseldorf, the BSPS conference in Manchester, the GAP.9 conference in Osnabrück, and the Karlsruhe Institute for Technology.

References

Adler, Caroline E. and Gertrude Hirsch Hadorn. 2014. The IPCC and treatment of uncertainties: topics and sources of dissensus. *WIRES Climate Change* 5: 663-676.

Aven, Terje and Ortwin Renn. 2015. An Evaluation of the Treatment of Risk and Uncertainties in the IPCC Reports on Climate Change. *Risk Analysis* 35: 701-712.

Budescu, David V., Han-Hui Por, Stephen B. Broomell, and Michael Smithson. 2014. The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*: 1-5. doi:10.1038/nclimate2194.

Carnap, Rudolf. 1947. On the Application of Inductive Logic. *Philosophy and Phenomenological Research* 8: 133-148.

Clarke, Brendan, Donald Gillies, Phyllis Illari, Frederica Russo, and Jon Williamson. 2014. Mechanisms and the Evidence Hierarchy. *Topoi* 33: 339-360.

Hollande, Francois. 2015. Ouverture du Leaders' event at annual conference of parties 21. https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/cop21cmp11_statement_hollande.pdf. Accessed: 5 April 2016.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of Intergovernmental Panel on Climate Change*. Eds. Stocker, Thomas F., Dahe Qin, Gian-Kasper Plattner, Melinda M. B. Tignor, Simon K. Allen, Judith Boschung, Alexander Nauels, Yu Xia, Vincent Bex, and Pauline M. Midgley. Cambridge, United Kingdom and New York: Cambridge University Press.

IPCC. 2014a. Intergovernmental panel on climate change: organization. http://www.ipcc.ch/organization/organization.shtml. Accessed: 5 April 2016.

IPCC. 2014b. Intergovernmental panel on climate change: working groups / task force. http://www.ipcc.ch/working_groups/working_groups.shtml. Accessed: 5 April 2016.

Jones, Roger N. 2011. The latest iteration of IPCC uncertainty guidance: an author perspective. *Climate Change* 108: 733-743.

Kelly, Thomas. 2005. The Epistemic Significance of Disagreement, in: Tamar S. Gendler and John Hawthorne (eds.) *Oxford Studies in Epistemology*. Oxford: Oxford University Press: 167-196.

Mastrandrea, Michael D., Christopher B. Field, Thomas F. Stocker, Ottmar Edenhofer, Kristie L. Ebi, David J. Frame, Hermann Held, Elmar Kriegler, Katharine J. Mach, Patrick R. Matschoss, Gian-Kasper Plattner, Gary W. Yohe, and Francis Zwiers. 2010. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change. Available at http://www.ipcc.ch. Accessed: 5 April 2016.

Mastrandrea, Michael D., Katharine J. Mach, Gian-Kasper Plattner, Ottmar Edenhofer, Thomas F. Stocker, Christopher B. Field, Kristie L. Ebi, and Patrick R. Matschoss. 2011. The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Climate Change* 108: 675-691.

Morgan, M. Granger. 2014. Use and (abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences of the United States of America* 111: 7176-7184.

Socolow, Robert H. 2011. High-consequence outcomes and internal disagreements: tell us more, please. *Climate Change* 108: 775-790.

Stegenga, Jacob. 2011. Is meta-analysis the platinum standard of evidence? *Studies in History and Philosophy of Biological and Biomedical Sciences* 42: 497-507.

United Nations. 2015. Conference of the Parties Twenty-first session Paris 30 November to 11 December 2015: Adoption of the Paris Agreement. http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf. Accessed: 5 April 2016.