

Richard Bradley, Casey Helgeson and Brian Hill
**Climate change assessments: confidence,
probability and decision**

**Article (Accepted version)
(Refereed)**

Original citation: Bradley, Richard and Helgeson, Casey and Hill, Brian (2017) *Climate change assessments: confidence, probability and decision*. *Philosophy of Science*, 84 (3). pp. 500-522.
ISSN 0031-8248
DOI: [10.1086/692145](https://doi.org/10.1086/692145)

© 2017 [Philosophy of Science Association](#)

This version available at: <http://eprints.lse.ac.uk/67373/>
Available in LSE Research Online: July 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Climate Change Assessments: Confidence, Probability and Decision*

Richard Bradley[†] Casey Helgeson[‡] Brian Hill[§]

July 30, 2016

Abstract. The Intergovernmental Panel on Climate Change has developed a novel framework for assessing and communicating uncertainty in the findings published in their periodic assessment reports. But how should these uncertainty assessments inform decisions? We take a formal decision-making perspective to investigate how scientific input formulated in the IPCC's novel framework might inform decisions in a principled way through a normative decision model.

1 Introduction

Assessment Reports produced by the Intergovernmental Panel on Climate Change (IPCC) periodically summarize the present state of knowledge about climate change, its impacts, and the prospects for mitigation and adaptation. More than 800 lead authors and review editors (and an even greater number of subsidiary authors and reviewers) contributed to the fifth and most recent report, which comprises a tome from each of three working groups,

*We would like to thank participants at the Centre for Philosophy of Natural and Social Science, the Grantham institute, the Centre for Analysis of Time Series (LSE), and at conferences in Helsinki (CLMPS), London (LSE) and Paris (Paris IV) for stimulating discussion and helpful feedback. Bradley and Helgeson gratefully acknowledge support from the AHRC grant Managing Severe Uncertainty (AH/J006033/1). Hill gratefully acknowledges support from the ANR DUSUCA (ANR-14-CE29-0003-01).

[†]Department of Philosophy, Logic and Scientific Method, LSE (R.Bradley@lse.ac.uk)

[‡]Centre for Philosophy of Natural and Social Science, LSE (C.Helgeson@lse.ac.uk)

[§]GREGHEC, CNRS, HEC Paris, Université Paris-Saclay (hill@hec.fr)

the condensed technical summaries, approachable summaries for policymakers, and a synthesis report. There is no *new* research in an IPCC report; the aim is rather to comprehensively assess existing research and report on the state of scientific knowledge. It is an unusually generous allotment of scientific resources to summary, review, and communication, reflecting the pressing need for authoritative scientific findings to inform policy-making in an atmosphere of skepticism and powerful status-quo interests.

Scientific knowledge comes in degrees of uncertainty, and the IPCC has developed an innovative approach to characterizing and communicating this uncertainty.¹ Their primary tools are probability intervals and a qualitative notion of *confidence*. In the reports' most carefully-framed findings, the two metrics are used together, with confidence assessments qualifying statements of probability. The question we examine here is how such findings might be incorporated into a normative decision framework. While the IPCC's treatment of uncertainties has been discussed extensively in the scientific literature and in a major external review (Shapiro et al., 2010), our question has not yet been addressed.

By exploring how scientific input in this novel format might systematically inform rational decisions, we hope ultimately to improve climate change decision-making and to make IPCC findings more useful to consumers of the reports. As will emerge below, the immediate lessons of this paper concern how the decision-theoretic perspective can help shape the IPCC's uncertainty framework itself, and how that framework is used by IPCC authors. One broader theoretical aim is to learn from the IPCC's experience with uncertainty assessment to improve evidence-based policy making more generally.

We begin by explaining the IPCC's approach to uncertainty in greater detail. We then survey recent work in decision theory that makes room for second-order uncertainty of the kind conveyed by IPCC confidence assessments when those assessments qualify probabilities. We identify a family of decision models (Hill, 2013) as particularly promising for our purposes, given IPCC practice and general features of the policy decision context. We show how to map IPCC-style findings onto these models, and, based on the resulting picture of how such findings inform decisions, we draw some lessons about the way the IPCC uncertainty framework is currently being used.

¹Unless specified, we use uncertainty in the everyday sense, rather than in the technical sense popular in economics, for instance.

2 Uncertainty in IPCC Reports

The fifth and most recent assessment report (AR5) affirms unequivocally that the earth’s climate system is warming and leaves little room for doubt that human activities are largely to blame. Yet climate change researchers continue to wrestle with deep and persistent uncertainties regarding many of the specifics, such as the pace of change in coming decades, the extent and distribution of impacts, or the prospect of passing potentially calamitous “tipping points.” Further research can, to a degree, reduce some of this uncertainty, but meanwhile it must be characterized, conveyed, and acted upon. Communication of uncertainty by IPCC authors is informed by an evolving set of *guidance notes* that share best practices and promote consistency across chapters and working groups (Moss and Schneider, 2000; Manning, 2005; Mastrandrea et al., 2010). These documents also anchor a growing, interdisciplinary literature devoted to the treatment of uncertainties within IPCC reports (Adler and Hadorn, 2014; Yohe and Oppenheimer, 2011).

One conspicuous feature of IPCC practice is the use of *confidence assessments* to convey a qualitative judgement about the extent of the evidence that backs up a given finding. Naturally, this varies from one finding to the next. And it is, intuitively, important information for policymakers. The format for expressing confidence has changed subtly from one IPCC cycle to the next, in part responding to critical review (Shapiro et al., 2010). Likewise for the *de facto* implementation within each working group (Mastrandrea and Mach, 2011). In AR5, confidence assessments are plentiful across all three working groups, from the exhaustive, unabridged reports through all of the summary and synthesis. The current guidance offers five qualifiers for expressing confidence: *very low*, *low*, *medium*, *high*, and *very high*. To pick the right one, an author team appraises two aspects of the evidence, (roughly): how much there is (considering quantity, quality and variety), and how well different sources of evidence agree. The more evidence, and the more agreement, the more confidence (Mastrandrea et al., 2010).

The second approved uncertainty metric is *probability*.² And by far the most common mode of presenting probabilities in AR5 is through words chosen from a preset menu of calibrated language, where, for example, *likely* has an official translation of “66-100% chance,” *virtually certain* means “99-100% chance,” and *more likely than not* means “>50% chance.” There are ten phrases on the menu, each indicating a different probability interval. (Precise probability density functions are also sanctioned where there is suf-

²The IPCC uses the term “likelihood,” though one should not read into this the technical meaning from statistics. We will use the more neutral term “probability.”

ficient evidence, though authors rarely exercise this option; percentiles from cumulative density functions are somewhat more common.)

Different author teams make somewhat different choices as they adapt the common framework to the particulars of their subject area. One way that authors have used the metrics is in combination: where the finding that is qualified by a confidence assessment is itself a probabilistic statement. In this case confidence is pushed into second position in a now two-stage characterization of overall uncertainty:

In the Northern Hemisphere, 1983–2012 was *likely* the warmest 30-year period of the last 1400 years (medium confidence).

Multiple lines of evidence provide high confidence that an [Equilibrium Climate Sensitivity] value less than 1°C is *extremely unlikely*.

Many, though not all, IPCC findings satisfy this format. Plenty of confidence assessments do something other than modify a probability claim, such as when an author team expresses confidence in an observational trend, or gives a blanket appraisal of projections from a given modeling approach. All probabilities, however, should be read as confidence-qualified. Sometimes the confidence level is not written out explicitly, but the guidance note instructs that “a finding that includes a probabilistic measure of uncertainty does not require explicit mention of the level of confidence associated with that finding if the level of confidence is ‘high’ or ‘very high’” (Mastrandrea et al., 2010, 3), meaning that readers should take unaccompanied probabilities to enjoy high or very high confidence. Findings reported in the form of the quotations above will be our focus here.

3 Decision, Imprecision and Confidence

Like any assessment that reflects a state of knowledge (or belief), the judgments of the IPCC can play two sorts of roles. On the one hand, they can represent the salient features of the world and our uncertainty about them; on the other hand, they can inform behavior, or policy. Any representation of uncertainty can be evaluated by its capacity to fulfill each of these roles. While the IPCC uncertainty framework has been developed mainly with the former role in mind—and we shall assume for the purposes of this paper that it fares sufficiently well on this front—the focus here is on the latter role. Are there existing normatively reasonable accounts of decision making into

which the IPCC representation of uncertainty provides relevant input, and what are the consequences of bringing the two together?

At first pass, the IPCC's uncertainty framework seems far-removed from models developed by decision theorists. The standard approach in decision theory, often termed *Bayesianism*, prescribes maximization of expected utility relative to the probabilities of the possible states of the world and the utilities of the possible consequences of available actions. Naturally, in order to apply this theory, the decision maker must be equipped with all decision-relevant probabilities. (Utilities are also required, but as they reflect judgements of value or desirability, they should come not from scientific reports but from society or the policy maker.) What the IPCC delivers, however, are not precise probabilities but *probability ranges*, qualified by *confidence* judgements. The former are too imprecise to be used in the standard expected utility model; the latter have no role at all to play in that model. IPCC findings thus sit uncomfortably with standard decision theory.

This mismatch need not reflect badly on the IPCC framework. On the contrary, several researchers (Bradley, 2009; Joyce, 2011; Gilboa et al., 2009, 2012; Gilboa and Marinacci, 2013) have suggested that the standard insistence on a single precise probability function leads to an inadequate representation of uncertainty, and may moreover have unintuitive, and indeed normatively undesirable consequences for decision. This has motivated the development, within both philosophy and economics, of alternative theories of rational decision making, on which we can draw in our attempt to accommodate scientific findings expressed using the IPCC uncertainty framework. This literature is large, and here we will consider only a few prominent models in our search for one with the appropriate features. And while at points some remarks on what constitutes reasonable decision behavior may be in order, we will not attempt a thorough normative comparison of models, instead referring the interested reader to the literature for discussion.

3.1 Imprecise Probability

The use of probability ranges by the IPCC invokes what is sometimes known in the theoretical literature as *imprecise probability*. Central to much of this literature is the use of sets of probability functions to represent the epistemic state of an agent who cannot determine a unique probability for all events of interest. Informally we can think of this set as containing those probability functions that the decision maker regards as permissible to adopt given the information she holds.

To motivate the idea, recall Popper's *paradox of ideal evidence* (1974, 407–8), which contrasts two cases in which we are asked to provide a probability

for a coin landing heads. In the first, we know nothing about the coin; in the second, we have already observed 1000 tosses and seen that it lands heads roughly half the time. Our epistemic state in the second case can reasonably be represented by a precise probability of one-half for the outcome of heads on the next toss. By contrast, the thought goes, the evidence available in the first case can justify only a set of probabilities—perhaps, indeed, the set of all possible probabilities. To adhere to a single number, even the “neutral” probability of one-half, would require a leap of faith from the decision maker, and it is hard to see why she should be forced to make this leap. Pragmatic considerations too suggest allowing imprecise probabilities. Given a choice between betting in the first case or in the second, it seems natural that one might prefer betting in the second—but a Bayesian decision maker cannot have such preferences.³

Bayesian decision insist that no matter the scarcity of the decision maker’s information, she must pick a single probability function to be used in all decisions. This is often called her “subjective” probability, and particularly in cases where the available information (combined with the decision maker’s expertise and personal judgement) provides little guidance, the “subjective” element may be rather hefty indeed. This probability function determines, together with a utility function on consequences, an expected utility for each action available to the decision-maker, and the theory enjoins her to choose the action with greatest expected utility.

Despite the severity of uncertainty faced in the climate domain, Bayesian decision theory has its adherents. John Broome, for instance, argues that in climate policy decision making:

The lack of firm probabilities is not a reason to give up expected value theory. You might despair and adopt some other way of coping with uncertainty . . . That would be a mistake. Stick with expected value theory, since it is very well founded, and do your best with probabilities and values. (Broome, 2012, 129)

Paralleling the points made in the coin example above, critics of the Bayesian view argue that the decision maker may be unable to supply the required precise subjective probabilities, and that any “filling in” of the gap between probability ranges and precise probabilities may prove too arbitrary to be a reasonable guide to decision. Policy makers may quite reasonably refuse to base a policy decision on a flimsy information base, especially when there is a lot at stake.

³The incompatibility of these and other reasonable preferences with Bayesianism is at the heart of the structurally similar Ellsberg paradox (Ellsberg, 1961).

Imprecise probabilists, on the other hand, face the problem of spelling out how a decision maker with a set of probability functions should choose. Her problem can be put in the following way. Each probability function in her set determines, together with a utility function on consequences, an expected utility for each available action; but except on rare occasions when one action *dominates* all others in the sense that its expected utility is greatest relative to *every* admissible probability function, this does not provide a sufficient basis for choice. Were the decision maker to simply average the expected utilities associated with each action, her decisions would then be indistinguishable from those of a Bayesian. There are, however, other considerations that she can bring to bare on the problem which will lead her to act in a way which cannot be given a Bayesian rationalisation. She may wish, for instance, to act cautiously, by giving more weight to the “down-side” risks (the possible negative consequences of an action) than the “up-side” chances, or by preferring actions with a narrower spread of (expected) outcomes.

A much-discussed decision rule encoding such caution is the maximin expected utility rule (MMEU), which recommends picking the action with the greatest minimum expected utility relative to the set of probabilities that the decision maker is working with (Gilboa and Schmeidler, 1989). To state the rule more formally, let $C = \{p_1, \dots, p_n\}$ be a set of probability functions,⁴ and for any $p \in C$ and action f , let $EU_p(f)$ be the expected utility of f computed from p . The rule then ascribes a value V to each action f in accordance with:

$$\text{(MMEU)} \quad V(f) = \min_{p \in C} [EU_p(f)]$$

MMEU is simple to use, but arguably too cautious, paying no attention at all to the full spread of possible expected utilities. This shortcoming is mitigated in some of the other rules for decision making that draw on imprecise probabilities, such as maximizing a weighted average of the minimum and maximum expected utility (often called the α -MEU rule), or of the minimum and mean expected utility, where the averaging weights can be thought of as reflecting either the decision maker’s pessimism or their degree of caution (see for instance Binmore, 2008; Ghirardato et al., 2004).

A question that all such rules must address is how to specify the set C of probabilities on which they are based. Where evidence is sparse, the Bayesian insistence on a single probability function seems too extreme, but if C contains all probabilities *logically consistent* with the evidence, then the decision maker is likely to end up with very wide probability intervals, which can in turn lead to overly cautious decision-making. A natural thought is

⁴For simplicity we suppose a finite set, but it needn’t be so.

that C should determine probability intervals only so broad as to ensure the decision-maker is confident the “true” probabilities lie within them, or that they contain all *reasonable* values (see, e.g., Gärdenfors and Sahlin, 1982). The decision maker may, for instance, wish to discard some implausible probability functions even though they are not, strictly speaking, contradicted by the evidence. Or if the source of these probabilities are the opinions of others, the decision maker need not consider every opinion consistent with the evidence, but rather only those in which they have some confidence. But how confident need they be? We return to this question below, after discussing the notion of confidence in more detail.

3.2 Confidence

The decision rules canvassed above can make use of the probability ranges found in IPCC reports, but not the confidence judgements that qualify them. According to the authors of the IPCC guidance notes “A level of confidence provides a qualitative synthesis of an author team’s judgment about the validity of a finding; it integrates the evaluation of evidence and agreement in one metric” (Mastrandrea et al., 2011, 679). Let’s address these two contributors to IPCC confidence in turn.

“Evaluation of evidence” depends on the amount, or weight, of the evidence relevant to the judgement in question. Suppose for instance that a decision maker is pressed to report a single number for the chance of heads on the next coin toss in the two cases described above, namely where she knows nothing about the coin and where she has already observed many tosses. She may report one-half in both cases, but is likely to have more confidence in that assessment in the case where the judgement is based on abundant evidence (the previously observed tosses). This is because a larger body of evidence is likely to be reflected in a higher level of confidence in the judgements that are based on it.

The second contributor to confidence is “agreement.” To tweak the coin example, compare a situation in which a group of coin experts agrees that the probability of heads on the next toss is one half with a second case where the same group is evenly split between those that think the probability is zero and those that think it is one. Here too, a decision maker pressed to give a single number might say one-half in both cases, but it seems reasonable to have more confidence in the first case than in the second. Holding the amount of evidence fixed, greater agreement engenders greater confidence.

The two dimensions of IPCC confidence connect to largely distinct literatures. The evidence dimension connects with that on the *weight of evidence* behind a probability judgement and how this weight can be included

in representations of uncertainty (Keynes, 1921). The agreement dimension connects with the literature on expert testimony and aggregation of probability functions (for a survey, see Genest and Zidek, 1986). Models employing confidence weights on different possible probabilities are to be found in both literatures. In the first, the probabilities are interpreted as different probabilistic hypotheses and the weights as measures of the agent's confidence in them. In the second, the probabilities are interpreted as the experts' judgements and the weights as a measure of an agent's confidence in the experts. So while weight of evidence and expert agreement are two distinct notions, they can be represented similarly, and play analogous roles in determining judgements and guiding action. It is thus not unreasonable to proceed in the manner suggested by the IPCC and place both under a single notion of confidence.

What role should these second-order confidence weights play in decision making? To the extent that different probability judgements support different assessments of the expected benefit or utility of an action, one would expect the relative confidence (or lack of it) that a decision maker might have in the former will transfer to the latter. It is then reasonable, *ceteris paribus*, to favor actions with high expected benefit based on the probabilities in which one has most confidence over actions whose case for being beneficial depends on probabilities in which one has less confidence.

One way to do this is to use the confidence weights over probability measures to weight the corresponding first-order expected utilities, determining what might be called the confidence-weighted expected utility (CWEU) of an action. Formally, let $C = \{p_1, \dots, p_n\}$ be a set of probability functions, and $\{\alpha_i\}$ the corresponding weights, normalised so that $\sum_i \alpha_i = 1$. Then:

$$\text{(CWEU)} \quad V(f) = \sum_i \alpha_i \cdot EU_{p_i}(f)$$

Here the weights effectively induce a second-order probability over C , and maximizing CWEU is equivalent to maximizing expected utility relative to the aggregated probability function obtained by averaging the elements of C using the second-order probabilities. But this seems unsatisfactory from a pragmatic point of view as it would preclude a decision maker displaying the sort of caution, or aversion to uncertainty, that we argued could be motivated in contexts like the coin example. Given a choice between betting on one coin or the other, an agent following CWEU cannot prefer betting on the coin for which she has more evidence. But some degree of discrimination between high and low confidence situations does seem appropriate for important policy decisions.

Other decision models proposed in the economics literature allow for this kind of discrimination. The “smooth ambiguity” model of Klibanoff et al. (2005) is a close variant of CWEU; it too uses second-order probability, but it allows for an aversion to wide spreads of expected utilities by valuing an action f in terms of the expectation (with respect to the second-order probability) of a transformation of the $EU_{p_i}(f)$, rather than the expected utilities themselves. Formally (and ignoring technicalities due to integration rather than summation), the rule works as follows:

$$\text{(KMM)} \quad V(f) = \sum_i \alpha_i \cdot \phi(EU_{p_i}(f))$$

where ϕ is a transformation of expected utilities capturing the decision maker’s attitudes to uncertainty (the decision maker displays aversion to uncertainty whenever ϕ is concave).

Other suggestions in this literature use general real-valued functions (rather than probabilities) at the second-order level, and can be thought of as refinements of the MMEU model discussed in the previous section. Gärdenfors and Sahlin (1982), for instance, uses the weights to determine the set of probability functions $C = \{p_1, \dots, p_n\}$ by admitting only those that exceed some confidence threshold; they then apply MMEU. Maccheroni et al. (2006) value each action as the minimum, across the set of probability functions C , of the sum of the action’s expected utility given p_i and the second-order weight given to p_i ; and Chateauneuf and Faro (2009) take the minimum of confidence-weighted expected utilities over probability functions with second-order weight beyond a certain absolute threshold.

From the perspective of this paper, all these proposals suffer from a fundamental limitation. Application of these rules requires a cardinal measure of confidence to serve as the weights on probability measures. That is, the numbers matter: otherwise it would make no sense to multiply or add them as is done by these rules. Whenever such confidence numbers are available, these models are applicable. The IPCC, however, commits to providing only a qualitative, ordinal measure of confidence: it can say whether there is more confidence or less, in one probability judgement compared to another, but not how much more or less.⁵ So the aforementioned models of decision require information of a different nature from what is furnished in IPCC reports.

IPCC practice is not unreasonable in this respect. Indeed if the decision maker has trouble forming precise first-order probabilities, why would he have any less trouble forming precise second-order confidence weights?

⁵Moreover, IPCC confidence applies to probability claims, not to (fully-specified) probability measures; it is not always straightforward to translate confidence in one to confidence in the other.

Such considerations plead in favor of a more parsimonious representation of confidence, in line with the ordinal ranking used by the IPCC. To connect this to decision making, however, a model is required that can work with ordinal confidence assessments without requiring cardinality. We have found only one such model in the literature, namely that proposed by Brian Hill (2013), which we now present in some detail.

3.3 Hill’s Decision Model

Hill’s central insight is that the probability judgements we adopt can reasonably vary with what is at stake in a decision. Consider, for instance, the schema for decision problems represented by Table 1, in which the option *Act* has a low probability of a very bad outcome (utility $x \ll 0$) and a high probability of a good outcome (utility $y > 0$). The table could represent a high stakes decision, such as whether to build a nuclear plant near a town when there is a small imprecise probability of an accident with catastrophic consequences. But it could equally well represent a low stakes situation in which the agent is deciding whether to get on the bus without a ticket when there is a small imprecise probability of being caught.

	prob. < 0.01	prob. $\geq .99$
Act	$x \ll 0$	$y > 0$
Don’t Act	0	0

Table 1: A small chance of a bad outcome.

Standard decision rules, such as expected utility maximisation or Maximin EU, are invariant with respect to the scaling of the utility function. Consequently they cannot treat a high stakes and a low stakes decision problem differently if outcomes in the former are simply a “magnification” of those in the latter—for instance if the nuclear accident was 100,000 times worse than being fined and the benefits of nuclear energy 100,000 times better than those of travelling for free. But it does not seem at all unreasonable to act more cautiously in high stakes situations than low stakes ones.⁶

To accommodate this intuition, Hill allows for the set of probability measures on which the decision is based to be shaped by how much is at stake in

⁶Some of the models mentioned in the previous section are also not invariant to scaling, though Hill (2013) argues that they do not properly capture the stakes dependence described below. In any case, their reliance on inputs that cannot be expressed in the IPCC’s confidence language is a reason to set aside these models for our purposes.

the decision. This stakes-sensitivity is mediated by confidence: each decision situation will determine an appropriate confidence level for decision making based on what is at stake in that decision. When the stakes are low, the decision maker may not need to have a great deal of confidence in a probability assessment in order to base her decision upon it. When the stakes are high, however, the decision maker will need a correspondingly high degree of confidence.

To formulate such a confidence-based decision rule, Hill’s model draws on a purely ordinal notion of confidence, requiring only that the set of probability measures forms a nested family of sets centred on the measures that represent the decision maker’s best-estimate probabilities. This structure is illustrated in Figure 1, where each circle is a set of probability measures. The inner-most set is assigned the lowest confidence level and each superset a higher confidence level than the one it encloses. These confidence assignments can be thought of as expressing the decision maker’s confidence that the “true” probability measure is contained in that set. Probability statements that hold for every measure in a superset enjoy greater confidence because the decision maker is more confident that the “true” measure endorses the statement.

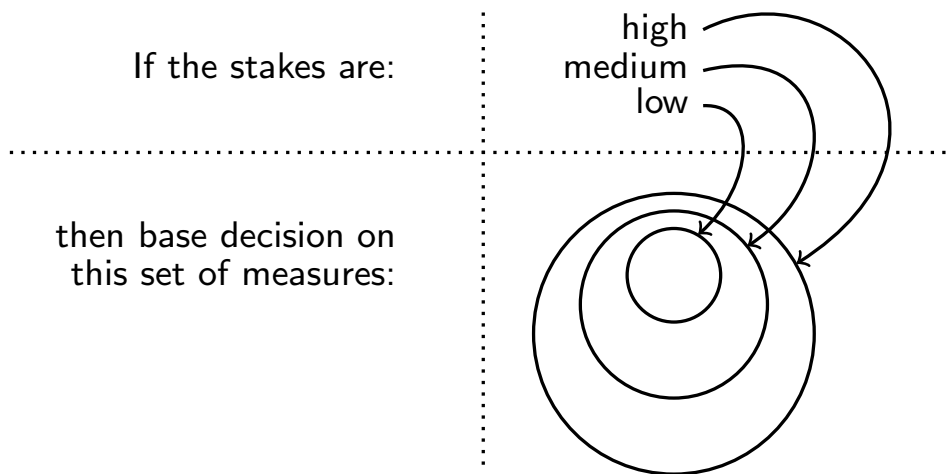


Figure 1: How much is at stake in a decision determines the set of probability measures that the decision rule can “see.”

For any given decision, the stakes determine the requisite level of confidence, which in turn determines the set of probability measures taken as the basis for choice: intuitively, the smallest set that enjoys the required level of

confidence.⁷ Once the set of measures has been picked out in this way, the decision maker can make use of one of the rules for decision making under ambiguity discussed earlier, such as MMEU or α -MEU. In the special case that the set picked out contains just one measure, ordinary expected utility maximisation is applicable.

As should be evident, what Hill provides is a schema for confidence-based decision rather than a specific model. Different notions of stakes will determine different confidence levels. And there is the question of which decision rule to apply in the final step. But these details are less important than the fact that the schema can incorporate roughly the kind of information that the IPCC provides. Spelling this out is our next task.

4 Confidence and IPCC Judgements

4.1 A Model of Confidence

Now we develop more formally the notion of confidence required to link IPCC communications to the model of decision making just introduced. As is standard, actions or policies will be modelled as functions from states of the world to outcomes, where outcomes are understood to pick out features of the world that the decision maker seeks to promote or inhibit. States are features of the world that, jointly with the actions, determine what outcome will eventuate. What counts as an outcome or a state depends on the context: when a decision concerns how to prepare for drought, for instance, mean temperatures may serve as states, while in the context of climate change mitigation they may serve as outcomes.

Central to our model is a distinction between two types of propositions that are the objects of different kinds of uncertainty: propositions concerning “ordinary” events, such as global mean surface temperature exceeding 21°C in 2050, and *probability propositions* such as there being a 50% chance that temperature will exceed 21°C in 2050. Intuitively the probability propositions represent possible judgements yielded by scientific models or by experts and, hence, are propositions in which the decision maker can have more or less confidence.

Let $S = \{s_1, s_2, \dots, s_n\}$ be a set of n states of the world and $\Omega = \{A, B, C, \dots\}$ be a Boolean algebra of sets of such states, called events or factual propositions. Let $\Pi = \{p_i\}$ be the set of all possible probability functions on Ω ,

⁷Formally, Hill (2013) requires both a measure of the stakes associated with a decision problem and a *cautiousness coefficient* which maps stakes onto confidence thresholds.

and $\Delta(\Pi)$ be the set of all subsets of Π .⁸ Members of $\Delta(\Pi)$ play a dual role: as both the possible imprecise belief states of the agent and as probability propositions, i.e., propositions about the probability of truth of the factual propositions in Ω . For instance, if X is the proposition that it will rain tomorrow, then the proposition that the probability of X is between one-half and three-quarters is given by the set of probability distributions p such that $0.5 \geq p(X) \geq 0.75$. So the probabilistic statements that are qualified by confidence assessments in the IPCC examples given in Section 2 correspond to elements of $\Delta(\Pi)$.

To represent the confidence assessments appearing in IPCC reports we introduce a weak pre-order, \succeq , on $\Delta(\Pi)$, i.e., a reflexive and transitive binary relation on sets of probability measures. Intuitively \succeq captures the relative confidence that a group of IPCC authors has in the various probability propositions about the state of the world, with $P_1 \succeq P_2$ meaning that they are at least as confident in the probability proposition expressed by P_1 as that expressed by P_2 , as would be the case if they gave P_2 a *medium confidence* assessment and P_1 *high confidence*. In practice, a confidence relation in line with IPCC findings will have up to five levels, corresponding to the five qualifiers in their confidence language (Section 2). It is reasonable to assume that \succeq is non-trivial (that $\Pi \succ \emptyset$) and monotonic with respect to logical implication between probability propositions (i.e., that $P_1 \succeq P_2$ whenever $P_2 \subseteq P_1$), because one should have more confidence in less precise propositions.

We do not, however, assume that \succeq is complete. Completeness can fail to hold for two different reasons. First, there may be issues represented in the state space about which the agent makes no confidence judgements. For example, the IPCC does not assess the chance of rain in London next week. Second, the agent may make a confidence judgement about a certain probability proposition, but no judgments about other probability propositions concerning the same issue. For example the IPCC may report medium confidence that a certain occurrence is *likely* (66–100% chance), but say nothing about how confident one should be that the same occurrence is *more likely than not* (50–100% chance).

Given cardinal confidence assessments of probability propositions, it is always possible to extract an underlying order \succeq summarizing the ordinal information. Hence this setup can also apply when cardinal information is available. But an order \succeq does not determine any cardinal confidence assessments. There will be a large family of such assessments, each consistent with

⁸We abstract from technicalities concerning the structure of the state space here and conduct the discussion as if everything were finite.

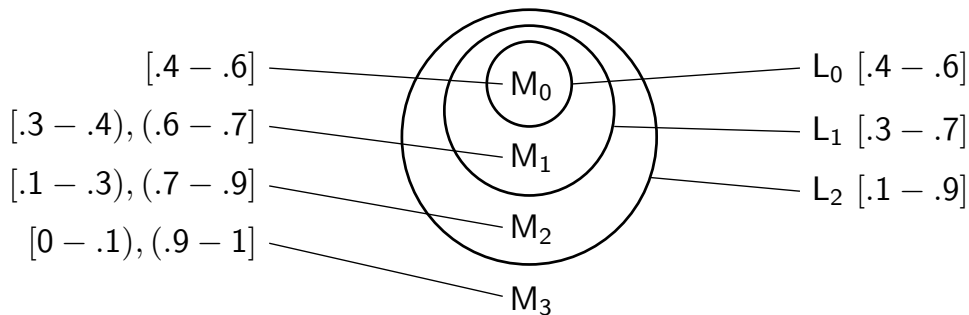
\succeq , which will not agree on any question concerning the numbers. The order thus drastically underspecifies the cardinal confidence measures required to apply the models discussed in Sect 3.2. By contrast, the ordinal information is precisely of the sort needed by Hill’s account.

Hill’s model of confidence effectively consists of a chain of probability propositions, $\{L_0, L_1, \dots, L_n\}$ with $L_i \subset L_{i+1}$. L_0 is the most precise probability proposition that the agent accepts; it summarizes her beliefs in the sense that every probability proposition that she accepts (with sufficient confidence) is implied by every probability function in L_0 . The other L_i are progressively less precise probability propositions held with progressively greater confidence. The chain $\{L_0, L_1, \dots, L_n\}$ is equivalent to what we shall call a *confidence partition*: an ordered partition of the space Π of probability measures. Any nested family of probabilities $\{L_0, \dots, L_n\}$ induces a confidence partition $\{M_0, \dots, M_n\}$ where $L_0 = M_0$ and $M_i = L_i - L_{i-1}$. M_i (for $i > 0$) contains those probability measures that the agent rules out as contenders for the “true” measures at the confidence level $i - 1$ but not at the higher confidence level i . Inversely, any confidence partition $\pi = \{M_0, \dots, M_n\}$ induces a nested family of sets of probability measures $\{L_0, \dots, L_n\}$ such that $L_0 = M_0$ and $L_i = M_0 \cup \dots \cup M_i$. A sample confidence partition and corresponding nested family are given in Figure 2, for the issue of the weather tomorrow. M_0 , the agent’s best-estimate probability range for rain, is the proposition that the probability of rain tomorrow is between 0.4 and 0.6, M_1 that it is either between 0.3 and 0.4 or between 0.6 and 0.7, M_2 that it is between 0.1 and 0.3 or 0.7 and 0.9, and M_3 the remaining probabilities. As is generally the case with the Hill model, the agent is represented by this figure as having made confidence judgements regarding any pair of these probability propositions for rain.

Which chain of probability propositions (or, equivalently, which confidence partition) does an IPCC-style assessment recommend for decision purposes? The probability measures in the lowest element of the partition are those that satisfy all of the probability propositions, on a given issue, that are affirmed by the IPCC (with any confidence level). The additional measures to be considered as contenders at the next level up, M_1 , need only satisfy those probability propositions affirmed by the IPCC with this next-level-up (or higher) level of confidence. Additional probability measures collected in M_2 should satisfy the IPCC probability propositions that are on or above the next confidence level up, and so on. Only confidence partitions satisfying these conditions faithfully capture the IPCC confidence and probability assessments.

Note that this protocol picks a unique confidence partition only in the case where the confidence relation \succeq is complete. Otherwise, several confidence

Figure 2: A confidence partition for the proposition that it will rain tomorrow. Bracketed intervals show probabilities given by the probability propositions. Nested sets L_0, L_1, L_3 can be constructed from the partition M_0, \dots, M_3 . The overall ordering is: $L_2 \supseteq L_1 \supseteq L_0 \equiv M_0 \supseteq M_1 \supseteq M_2 \supseteq M_3$.



partitions will be consistent with the confidence relation; as noted above, this will generally be the case for IPCC assessments. Since each confidence partition corresponds to a unique complete confidence relation, the use of a particular partition essentially amounts to “filling in” confidence assessments that were not provided. How best to “fill in” is an important question that is beyond the scope of this investigation.⁹ (In the final section below, we approach the issue from another perspective: how IPCC authors might reduce the need to perform this completion by providing more confidence assessments.)

Next we illustrate the confidence partition concept by applying it to a concrete example from the IPCC’s Fifth Assessment Report (AR5).

4.2 An Example

Equilibrium climate sensitivity (ECS) is often used as a single-number proxy for the overall behavior of the climate system in response to increasing greenhouse gas concentrations in the atmosphere. The greater the value, the greater the tendency to warm in response to greenhouse gasses. The quantity is defined by a hypothetical global experiment: start with the pre-industrial

⁹While we don’t necessarily recommend doing so, we note that from a purely formal perspective it is easy to define a unique “canonical” confidence partition for any incomplete relation: it suffices to take each M_i to be the largest set satisfying the conditions laid out above. This is analogous to the so-called “natural extension” in the imprecise probability literature (Walley, 1991). Similarly but perhaps more practically, one could use the largest set of probability measures from a certain family of distributions.

atmosphere and instantaneously double the concentration of carbon dioxide; now sit back and allow the system to reach its new equilibrium (this would take hundreds to thousands of years). ECS is the difference between the annual global mean surface temperature of the pre-industrial world and that of the new equilibrium world. In short, it answers the question: How much does the world warm if we double CO_2 ?

The most recent IPCC findings on ECS draw on several chapters of the Working Group One contribution to the AR5. Estimates of ECS are based on statistical analyses of the warming observed so far, similar analyses using simple to intermediate complexity climate models, reconstructions of climate change in the distant past (paleoclimate), as well as the behavior of the most complex, supercomputer-driven climate models used in the last two phases of the colossal Coupled Model Intercomparison Project (CMIP3 and CMIP5). An expert author team reviewed all of this research, weighing its strengths, weaknesses, and uncertainties, and came to the following collective judgements. With high confidence, ECS is *likely* in the range 1.5°C to 4.5°C and *extremely unlikely* less than 1°C . With medium confidence it is *very unlikely* greater than 6°C (Stocker et al., 2013, 81).

In light of the confidence model discussed above, reports of this kind can be understood in terms of a confidence partition over probability density functions (pdfs). Beginning from all possible pdfs on the real line—each one expressing a (precise) probability claim about ECS—think of what the author team is doing, as they evaluate and debate the evidence, as sorting those pdfs into a partition $\pi = \{M_0, \dots, M_n\}$. The findings cited above then communicate aspects of this confidence partition. To illustrate, we present a toy partition that exemplifies the IPCC’s findings on ECS.

Suppose the confidence partition has four elements $\pi = \{M_0, \dots, M_3\}$. For concreteness, we suppose that each element contains only lognormal distributions. Figure 3 displays the pdfs in the first two elements of the partition. The solid lines are the functions from M_0 ; collectively, these indicate what the IPCC’s experts regard as a plausible range of probabilities for ECS in light of the available evidence. The dashed lines are the pdfs in M_1 , which collectively represent a second tier of plausibility. M_2 is another step down from there, and M_3 is the bottom of the barrel—all of the pdfs more or less ruled out by the body of research that the experts evaluated. (M_2 and M_3 are not represented in Figure 3.)

Recall that the partition π generates a nested family of subsets $\{L_1, \dots, L_n\}$, where L_i is the union of P_0 through P_i and each L is associated with a level of confidence. Here we are concerned mainly with $L_0 = M_0$ and $L_1 = M_0 \cup M_1$, and we suppose in this case that L_0 corresponds to *medium confidence*, and L_1 to *high confidence*. To see how an IPCC-style finding follows from

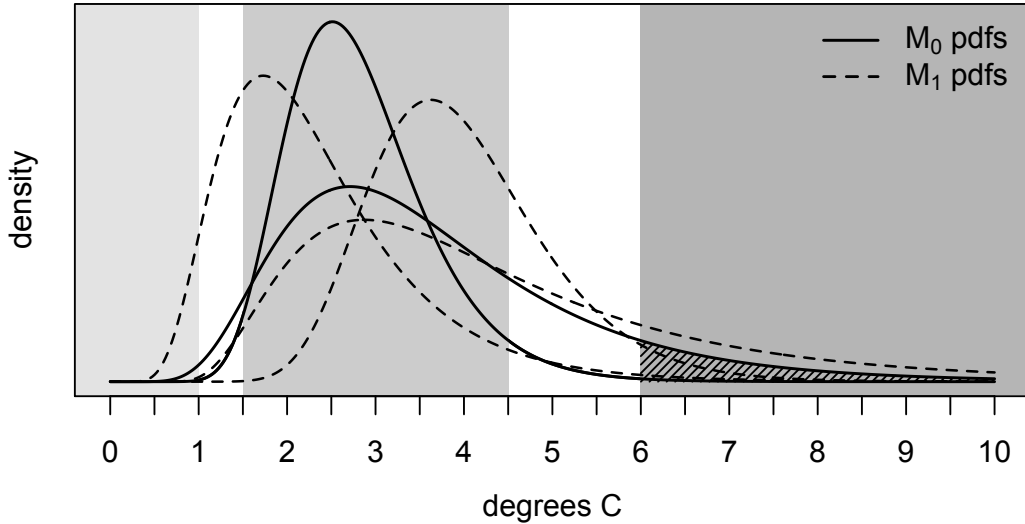


Figure 3: An illustration of a confidence partition consistent with IPCC findings on Equilibrium Climate Sensitivity. The hatched area corresponds to the finding that ECS is *very unlikely* greater than 6°C (medium confidence).

the confidence partition, consider what our partition says about ECS values above 6° . If we restrict attention to L_0 , there are only two pdfs to examine; one assigns (nearly) zero probability to values above 6° while the other assigns just under 0.1 probability (the shaded area in Figure 3). In the IPCC’s calibrated language, the probability range $0 - 0.1$ is called *very unlikely*, thus the finding: ECS is *very unlikely* greater than 6°C (medium confidence).

The IPCC’s two other findings on ECS are reflected in our partition as follows. Reporting with high confidence means broadening our view from L_0 to L_1 , taking the M_1 pdfs into account in addition to those in M_0 . The interval $1.5 - 4.5$ is indicated in Figure 3 in gray. The smallest probability given to that interval by any of the functions pictured is a little more than 0.6, and the highest probability is nearly 1, an interval that corresponds roughly with the meaning of *likely* ($0.66 - 1$). Regarding ECS values below 1, several pdfs give that region zero probability, while the most left-leaning of them gives it 0.05. The range $0 - 0.05$ is called *extremely unlikely*. Thus we have: ECS is *likely* in the range 1.5°C to 4.5°C and *extremely unlikely* less than 1°C (both with high confidence).

The three findings discussed above are far from the only ones that follow from the example partition. To report again on ECS values above 6°C , only now with high confidence rather than medium, the probability interval should

be expanded from $0 - 0.1$ to $0 - 0.2$ (0.2 being the area to the right of 6° under the fattest-tailed of the M_1 pdfs). Or to report on values below 1°C with medium confidence rather than high, the probability interval should be shrunk from $0 - 0.5$ to $0 - 0.01$ (*exceptionally unlikely*). The confidence partition determines an imprecise probability at medium confidence, and another at high confidence, for any interval of values for ECS.

It should be emphasized that these additional findings do not necessarily follow from the three that the IPCC in fact published. They follow from this particular confidence partition, which is constrained—though not fully determined—by the IPCC’s published findings.¹⁰ Asking what could be reported about ECS at *very high* confidence further highlights the limits of what the IPCC has conveyed. Suppose the set L_2 corresponds to *very high* confidence. As the IPCC has said nothing with *very high* confidence, we have no information about the pdfs that should go into M_2 , and thus L_2 , so we have no indication of how much the reported probability ranges should be expanded in order to claim very high confidence. This may be because in the confidence partition representation of the experts’ group beliefs, M_2 is a sprawling menagerie of pdfs. In this case probabilities at the very high confidence level would be so imprecise as to appear uninformative. On the other hand, it may sometimes be of interest to policymakers just how much (or how little) can be said at the very high confidence level.

5 Discussion

We now treat some possible objections, identify open questions and challenges, and point out some potential consequences of this decision-theoretic take on IPCC assessments.

Our model should be understood as illustrating how, in principle, such findings can be used in decision making. It provides a disciplining structure for the uncertainty expressed in IPCC findings—structure that is a prerequisite for the use of such findings within a normative decision model. (As noted above, there remains a gap between IPCC findings and the decision model in so far as the model involves a full confidence partition whereas the statements provided by the IPCC constrain but do not fully determine one.) Our model sketches one way in which such findings can be harnessed to provide concrete decision support, but other procedures for generating

¹⁰Regarding our extrapolated high-confidence finding on ECS exceeding 6°C , probability 0.2 does appear to be the upper limit under the restriction to lognormal distributions, though not in the absence of this restriction. This may be relevant to the issue of partition completion discussed in the previous section.

confidence partitions, or even for using the partial information without introducing new structure, deserve exploration. We have concentrated on the only decision model we could find that uses an ordinal confidence structure over probability statements, but it may be possible to develop alternative models with these features.

An interesting question concerns the implications of the model for judgements of joint probability and confidence. Suppose that two IPCC author groups respectively report with high confidence that low rainfall is *likely* and that low temperature is *likely*. What can be inferred about the prospect of both low rainfall *and* low temperature? This question turns on at least three issues. The first is the standard issue of joint probabilities. As is well known, one cannot conclude from these probability assessments that low rainfall and temperature is *likely*. And indeed, under the model set out in Section 4, nothing more than what follows from the individual probability propositions is assumed about the joint probability.

The second is the issue of “joint confidence.” The nested sets representation of confidence employed in Hill’s model implies that if both “low rainfall is *likely*” and “low temperature is *likely*” are held with high confidence, then their conjunction “low rainfall is *likely* and low temperature is *likely*” must be held with high confidence as well. This follows from the fact that a proposition is held with high confidence if it is supported by every probability function contained in the high-confidence set. On the other hand, it does not follow that the proposition “A combination of low rainfall and low temperature is *likely*” is held with high confidence since the high probability of this combination does not follow from high probability of its elements.

The third issue involves the *calibration* of confidence levels between groups. How do we know that what one group means by “high confidence” is the same as the other (and, indeed, that they mean the same thing to the policy maker using their findings)? A proper calibration scale would enable clear and unambiguous formulation and communication of confidence judgements across authors and actors. Were one to take our proposal for connecting the IPCC uncertainty language with theories of decision seriously, one major challenge is to develop such a scale. This development would likely go hand-in-hand with elicitation mechanisms that would allow IPCC authors to reveal and express their confidence in probability assessments.

Turning now to the *use* of the confidence partition in decision making, the Hill (2013) family of models gives confidence a role in pointing decision makers to the set of probability measures that is right for them in a given context. The decision maker’s utilities determine the stakes, and their cautiousness coefficient maps the stakes to a level of confidence and thus to the set of probability measures that their decision rule will take into account in

evaluating actions. IPCC findings inform the confidence element of Hill’s model, but they deliver neither a measure of the stakes associated with a decision problem nor a cautiousness coefficient. Where an individual acts alone, the stakes are determined by her preferences (or her utility function) while the cautiousness coefficient reflects some feature of her attitudes to uncertainty. In the case of climate policy decisions, things are analogous but more complicated. Putting utilities on outcomes and fixing the level of cautiousness are difficult tasks, insofar as both should reflect the interests and attitudes of individuals living in different places and at different times. That IPCC findings (at least those addressing the physical science basis of climate change) do not provide these elements is as it should be: this is not a “fact” dimension, on which climate scientists have expertise, but a “value” dimension, which derives from the stakeholders to the decision.

This fact-value distinction (or belief-taste distinction in economics) is muddled by many of the decision models surveyed above, for instance the MMEU decision model (Gilboa and Schmeidler, 1989), as well as those of Maccheroni et al. (2006) and Chateauneuf and Faro (2009), none of which permit a clean separation of beliefs from tastes.¹¹ In the case of MMEU for example, the set of probability functions captures both the beliefs or information at the decision maker’s disposal, and his taste for choosing in the face of uncertainty. Using a smaller range of probabilities can be interpreted as having a less cautious attitude towards one’s ignorance. Such models are less suitable in a policy decision context where scientists’ input should in principle be restricted to the domain of facts (and uncertainty about them).

In this context, it is another virtue of the Hill model that it does support a clean fact-value distinction (Hill, 2013, Sect 3). Confidence is exclusively a belief aspect, whereas the cautiousness coefficient is a taste factor. So the encroachment of value judgements into scientific reporting is not, at least, a theoretical consequence of the model.

This normatively attractive property of the model appears to be quite rare. Moreover, of the models that we have found in the literature which can incorporate something akin to the IPCC’s confidence and which can claim to support a neat belief-taste distinction, all require cardinal confidence assessments. This has already been noted for the most prominent such model, the smooth ambiguity model (Section 3.2). So, at least as far as we are aware, Hill’s model is the only available decision-theoretically solid

¹¹See Klibanoff et al. (2005, Sect 1, 3 & 5.1) for a clear discussion, or alternatively Gilboa and Marinacci (2013, Sect 3.5 & 4) and Hill (2016, Sect 3). This issue largely turns on so-called comparative statics results, reported in Gilboa and Marinacci (2013, Sect 4) for the MMEU model, and in the respective papers (Proposition 8 in each case) for the others.

representation that can capture the role of uncertainty about probability judgements without demanding value judgements from scientists or cardinal second-order confidence assessments. As such, our investigation provides a perhaps unexpected vindication of IPCC practice via the affinity between their uncertainty guidance and one of the only decision models that seems suitable for the climate policy decisions they aim to inform.

6 Recommendations

Our discussion of the IPCC’s uncertainty framework and the relevant policy decision requirements allows us to make several tentative recommendations.

In the climate sensitivity example above, we saw multiple statements each addressing a different range (left tail, middle, right tail) of the same uncertain quantity. And these statements used different confidence levels. But what we do not see in this example, nor have we found elsewhere, is multiple statements at different confidence levels, concerning the same range of the uncertain quantity. That is, we do not see pairs of claims such as: the chance that ECS exceeds 6° is, with *medium confidence*, less than 10%, and with *high confidence* less than 20%. The confidence partition formalism shows how it can make sense, conceptually, to answer the same question at multiple confidence levels. Doing so gives a richer picture of scientific knowledge, and the added information may be valuable to policy makers and to the public. There is no basis for the current (unwritten) convention of reporting only a single confidence level; a richer reporting practice is possible, and appears desirable.

Given the possibility of reporting at more than one confidence level for a given finding, in choosing just one, IPCC authors are implicitly managing a trade-off between the size of a probability interval and the level of confidence (e.g., *likely* (.66 – 1) with *medium confidence*, versus *more likely than not* (.5 – 1) with *high confidence*). Yet the uncertainty guidance notes offer no advice to authors on managing this trade-off.¹² Moreover, in light of the decision model developed above, there is an aspect to this choice that falls on the value side of the fact-value divide. While in practice IPCC authors may select on epistemic grounds (where they can make the most informative statements), the choice may be understood as involving a value judgement, since it may appear to suggest which set of probability measures the reader should use to decide. Normally it is the agent’s utilities and cautiousness that

¹²The AR4 guidance note included the advice to “Avoid trivializing statements just to increase their confidence” (Manning, 2005, 1). Note, however, that the meaning of “confidence” changed between AR4 and AR5.

together pick out the appropriate set from the nested family of probability measures. So not only is reporting at multiple confidence levels conceptually sensible, but it may be desirable in order simultaneously to give relevant information to different users who will determine for themselves the level of confidence at which they require probabilistic information to inform their decisions.

Naturally, it is impractical to demand that IPCC reports provide assessments at every confidence level on every issue that they treat probabilistically. But a feasible step in that direction might be to encourage reporting at more than one level, where the evidence allows, and when the results would be informative. The value-judgment aspect of confidence suggests a second step. The choice of confidence level(s) at which IPCC authors assess probability would ideally be informed in some way by the public or its representatives, suggesting that policy makers should be involved at the beginning of the IPCC process, to provide input regarding the confidence level(s) at which scientific assessments would be most decision-relevant. Communication of the relevant confidence level between policy makers and climate scientists would rely on and be formulated in terms of the sort of calibration scale touched on above. There are, of course, many decisions to be taken, with different stakes and stakeholders: mitigation decisions and adaptation decisions, public and private, global, regional, and local. The envisioned policy maker and stakeholder input would presumably indicate varying levels of confidence for key findings across IPCC chapters and working groups.

The realm of recommendations and possibilities goes well beyond those explored here. Our aim is simply to suggest some ideas for guiding practice on the basis of how IPCC assessments can be used in decision and policy making and, more importantly, to open a discussion on the issue.

References

- Adler, C. E. and G. H. Hadorn (2014). The IPCC and treatment of uncertainties: Topics and sources of dissensus. *Wiley Interdisciplinary Reviews: Climate Change*.
- Binmore, K. (2008). *Rational decisions*. Princeton University Press.
- Bradley, R. (2009). Revising incomplete attitudes. *Synthese* 171(2), 235–256.
- Broome, J. (2012). *Climate Matters: Ethics in a Warming World (Norton Global Ethics Series)*. WW Norton & Company.
- Chateauneuf, A. and J. H. Faro (2009). Ambiguity through confidence functions. *Journal of Mathematical Economics* 45(9), 535–558.

- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics* (75), 643–669.
- Gärdenfors, P. and N.-E. Sahlin (1982). Unreliable probabilities, risk taking, and decision making. *Synthese* 53(3), 361–386.
- Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 114–135.
- Ghirardato, P., F. Maccheroni, and M. Marinacci (2004). Differentiating ambiguity and ambiguity attitude. *J. Econ. Theory* 118(2), 133–173.
- Gilboa, I. and M. Marinacci (2013). Ambiguity and the Bayesian paradigm. In *Advances in economics and econometrics, tenth world congress*, Volume 1.
- Gilboa, I., A. Postlewaite, and D. Schmeidler (2009). Is it always rational to satisfy Savage’s axioms? *Economics and Philosophy*, 285–296.
- Gilboa, I., A. Postlewaite, and D. Schmeidler (2012). Rationality of belief or: why savage’s axioms are neither necessary nor sufficient for rationality. *Synthese*.
- Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics* 18(2), 141–153.
- Hill, B. (2013). Confidence and decision. *Games and Economic Behavior* 82, 675–692.
- Hill, B. (2016). Confidence in Beliefs and Rational Decision Making. mimeo, HEC Paris.
- Joyce, J. M. (2011). A defense of imprecise credences in inference and decision making. *Oxford Studies in Epistemology* 4.
- Keynes, J. M. (1921). A treatise on probability. Vol. 8 of *Collected Writings* (1973 ed.).
- Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica* 73(6), 1849–1892.
- Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 74(6), 1447–1498.
- Manning, M. R. (2005). Guidance notes for lead authors of the IPCC fourth assessment report on addressing uncertainties. Technical report, Intergovernmental Panel on Climate Change (IPCC).

- Mastrandrea, M. D., C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, E. Kriegler, K. J. Mach, P. R. Matschoss, G.-K. Plattner, G. W. Yohe, and F. W. . . Zwiers (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Technical report, Intergovernmental Panel on Climate Change (IPCC). Available at <<http://www.ipcc.ch>>.
- Mastrandrea, M. D. and K. J. Mach (2011). Treatment of uncertainties in IPCC assessment reports: past approaches and considerations for the Fifth Assessment Report. *Climatic Change* 108(4), 659–673.
- Mastrandrea, M. D., K. J. Mach, G.-K. Plattner, O. Edenhofer, T. F. Stocker, C. B. Field, K. L. Ebi, and P. R. Matschoss (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change* 108(4), 675–691.
- Moss, R. H. and S. H. Schneider (2000). Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In R. Pachauri, T. Taniguchi, and K. Tanaka (Eds.), *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*, IPCC Supporting Material. Intergovernmental Panel on Climate Change (IPCC).
- Popper, K. R. (1974). *The Logic of Scientific Discovery* (6 ed.). London: Hutchinson.
- Shapiro, H. T., R. Diab, C. de Brito Cruz, M. Cropper, J. Fang, L. Fresco, S. Manabe, G. Mehta, M. Molina, P. Williams, et al. (2010). Climate change assessments: Review of the processes and procedures of the IPCC. Technical report, InterAcademy Council, Amsterdam.
- Stocker, T., D. Qin, G.-K. Plattner, L. Alexander, S. Allen, N. Bindoff, F.-M. Bréon, J. Church, U. Cubasch, S. Emori, P. Forster, P. Friedlingstein, N. Gillett, J. Gregory, D. Hartmann, E. Jansen, B. Kirtman, R. Knutti, K. K. Kumar, P. Lemke, J. Marotzke, V. Masson-Delmotte, G. Meehl, I. Mokhov, S. Piao, V. Ramaswamy, D. Randall, M. Rhein, M. Rojas, C. Sabine, D. Shindell, L. Talley, D. Vaughan, and S.-P. Xie (2013). Technical summary. In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (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*, pp. 33–115. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*, Volume 42. London: Chapman & Hall.

Yohe, G. and M. Oppenheimer (2011). Evaluation, characterization, and communication of uncertainty by the intergovernmental panel on climate change—an introductory essay. *Climatic Change* 108(4), 629–639.