PHILOSOPHICAL TRANSACTIONS A

False Precision, Surprise and Improved Uncertainty Assessment

Journal:	*Philosophical Transactions A*
Manuscript ID:	RSTA-2014-0453.R1
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Parker, Wendy; Durham University, Philosophy Risbey, James; CSIRO, Oceans and Atmosphere
Issue Code: Click here to find the code for your issue.:	RESP-CC
Subject:	Climatology < EARTH SCIENCES
Keywords:	uncertainty, surprise, climate change, models, probability, unknown unknowns

SCHOLARONE™
Manuscripts

Page 1 of 14

Submitted to Phil. Trans. R. Soc. A - Issue

PHILOSOPHICAL TRANSACTIONS A

False Precision, Surprise and Improved Uncertainty Assessment

Wendy S. Parker[1*] and James S. Risbey[2]

[1]*Department of Philosophy, Durham University, Durham UK*
[2]*CSIRO Oceans and Atmosphere, Hobart, Tasmania*

An uncertainty report describes the extent of an agent's uncertainty about some matter. We identify two basic requirements for uncertainty reports, which we call *faithfulness* and *completeness*. We then discuss two pitfalls of uncertainty assessment that often result in reports that fail to meet these requirements. The first involves adopting a one-size-fits-all approach to the representation of uncertainty, while the second involves failing to take account of the risk of surprises. In connection with the latter, we respond to the objection that it is impossible to account for the risk of genuine surprises. After outlining some steps that both scientists and the bodies who commission uncertainty assessments can take to help avoid these pitfalls, we explain why striving for faithfulness and completeness is important.

Keywords: *uncertainty, surprise, climate change, models, probability, unknown unknowns*

1. Introduction

Many questions of interest to decision makers are empirical questions that science can help to answer. Do levels of air pollution in our region regularly exceed target levels? What causes such elevated pollution levels? What are the health consequences? Though answers to empirical questions like these are never logically certain, in some cases the uncertainty is negligible; the answers are beyond any reasonable doubt and can be reported in a definitive way. For instance, the evidence regarding the levels of air pollution in our region may be so strong that scientists can simply report: the levels regularly exceed the target levels. In the climate context too, definitive answers are given for some questions. The most recent Intergovernmental Panel on Climate Change (IPCC) report, for instance, concludes that "warming of the climate system is unequivocal" [1, p.4].

But for many empirical questions of interest to decisions makers, answers have non-negligible uncertainty. In the climate context, this is the case for many questions. What fraction of global warming in the period 1951-2010 was human caused? How would the frequency of droughts and floods in different regions change by the end of the 21st century under the A1B emission scenario? Available science does not provide definitive, single-valued answers to questions like these. Thus, the most recent IPCC report states instead that it is extremely likely (i.e. probability ≥0.95) that more than half of the late 20th century warming was human caused [1, p.17]. Uncertainty is indicated here both by reporting a range of values ("more than half") rather than a single value and by reporting that it is only "extremely likely" that the true value falls in that range.

*Author for correspondence (wendy.parker@durham.ac.uk).

This paper is concerned with good practice in assessing and reporting uncertainty.[1] Section 2 identifies two basic requirements for uncertainty reports. A *faithful* uncertainty report is one that accurately describes what an agent believes the extent of current uncertainty to be, while a *complete* uncertainty report is one that takes account of all significant sources of uncertainty and all available evidence. Section 3 discusses two common pitfalls of uncertainty assessment that can result in reports that fail to meet these basic requirements in even an approximate way. The first pitfall involves adopting a one-size-fits-all approach to the representation of uncertainty, while the second involves ignoring the risk of surprises. In connection with the latter, Section 4 responds to the objection that it is impossible to account for the risk of genuine surprises. Section 5 discusses some steps that scientists could take to help avoid these pitfalls and to improve uncertainty assessment more generally. Section 6 identifies several ways that governmental and industrial bodies that commission uncertainty assessments could also help in this regard. In closing, Section 7 reviews why striving to meet the basic requirements of faithfulness and completeness is important.

2. Two basic requirements for uncertainty reports

An uncertainty report describes the extent of an agent's uncertainty about some matter in light of available information.[2] When scientists are tasked with assessing and reporting uncertainty, this is typically understood to mean their uncertainty in light of information that is "accepted by the scientific community." The latter in turn is sometimes defined operationally to include only results that are published in the peer-reviewed scientific literature and/or produced via particular methodologies (e.g. via randomized controlled trials in medical contexts).

Uncertainty reports can take various forms and are sometimes described as having different "levels of precision" [2-3]. For example, a report of uncertainty about the future value of a climate variable, X, might come in the form of a statement that: (a) gives a full probability density function / probability distribution[3] over values of X; (b) gives a range of values of X in which the future value can be expected to fall with a precisely-specified probability, such as 0.95; (c) gives a range of values of X in which the future value can be expected to fall with an imprecise or interval probability, such as 0.6 – 0.9, or with a qualitative level of confidence, e.g. medium; (d) gives a range of values of X that can be considered plausible but indicates that probabilities cannot be assigned; (e) gives an order of magnitude estimate of the future value of X but indicates that more precise estimates are out of reach; (f) indicates that the future value of X will be greater than (or less than) the current value, though by how much is unclear; (g) admits that almost nothing is known about the future value of X.[4] Other forms of uncertainty report are possible as well.

Reports of uncertainty are often produced for the purpose of aiding decision making, including high-consequence decisions in government and industry. Here we discuss two basic requirements that such uncertainty reports should meet, which we refer to as *faithfulness* and *completeness*. Clearly there can

---

[1] Though most of the discussion applies to uncertainty assessment in general, we are particularly concerned with uncertainty in the context of climate change, which is often assessed in part with the help of climate models.

[2] Here, 'information' should be understood broadly; it can include theoretical understanding, observational data, modeling results, beliefs about the limitations of these information sources, and other background beliefs. What it means for information to be 'available' is less clear, but we assume that available information includes not only information that the agent (i) is consciously aware of or (ii) can easily retrieve from memory, but also basic implications of (i) and (ii) that the agent would recognize if she made even a relatively modest effort to do so.

[3] Note that the probabilities that appear in reports of type (a)-(c) are subjective probabilities, i.e. degrees of belief/confidence or credences. If an agent's degrees of belief are calibrated, then $(p*100)\%$ of the statements to which she assigns a probability of p turn out to be true statements. See also [4].

[4] Risbey and Kandlikar [2] refer to category (g) as "effective ignorance". Risbey and O'Kane [5] advocate greater openness to using this category (as needed) when reporting uncertainty, noting that it was not included among the levels of precision presented as options in the most recent IPCC guidance on uncertainty assessment [3].

3

be other desiderata for uncertainty reports as well, but faithfulness and completeness seem to be among the fundamental desiderata; in ordinary circumstances, when these requirements are not at least approximately met, an uncertainty report is inadequate.

According to the *faithfulness* requirement, an uncertainty report should accurately describe what the agent believes the extent of current uncertainty to be; it shouldn't imply that uncertainty is greater than, less than or otherwise different from what the agent actually believes it to be.[5] For example, if an agent concludes that available information indicates that X is more likely than not, then a report that "X is unlikely" would not meet the faithfulness requirement, nor would a report that "X is possible (but nothing more can be concluded)". The faithfulness requirement follows from the simple fact that, in ordinary circumstances, a request for an uncertainty report *just is* a request to know the agent's informed judgment about the extent of current uncertainty.

According to the *completeness* requirement, an uncertainty report should take account of all significant sources of uncertainty, and should consider all available (relevant) information when doing so. For example, suppose a scientist is tasked with estimating the uncertainty associated with model-projected changes in temperature in a region, and he recognizes that this uncertainty depends on three underlying sources of uncertainty, each of which is expected to make a significant contribution to the total uncertainty: initial condition uncertainty, parameter uncertainty and structural uncertainty.[6] If his analysis only takes into account initial condition uncertainty (e.g. perhaps he does an ensemble experiment using different initial conditions), then the analysis – and the uncertainty report it delivers – will be incomplete. Like the faithfulness requirement, the completeness requirement seems to follow from the very nature of the request for an uncertainty report: it is, in ordinary circumstances anyway, a request for a report of total uncertainty, based on all of the available evidence.[7]

Even with apparently straightforward requirements like these, there are complications. Here we discuss just two. First, as noted above, sometimes experts are asked to assess uncertainty in light of only a subset of the information available to them, e.g. in light of findings that meet a specified criterion. In that case, experts might try to meet the faithfulness and completeness conditions while assuming that only the specified subset of information is available, or they might try to convince those making the request that it would be better to consider additional available information. The latter seems the better option if there is reason to think that the additional information would make a significant difference to the uncertainty estimate produced.

Second, uncertainty assessments often are carried out by groups of experts, rather than lone individuals. Group assessment has the advantage that it allows for pooling of knowledge and expertise. A problem, however, is that even after open-minded discussion and reflection, individual group members sometimes have conflicting evaluations of the evidential significance of particular findings and reach different conclusions about the extent of current uncertainty. How then should "the group's" uncertainty be faithfully reported? Is there such thing as "the group's" uncertainty in this situation? This is a controversial matter, closely related to concerns about reporting only 'consensus' conclusions in expert assessment (see also [7-9]). Given space limitations, we focus on

---

[5] This is similar to what Parker [6] referred to as 'ownership' of an uncertainty report.

[6] This example is concerned with uncertainty about the changes in climate that would occur under a particular emission scenario. There is of course substantial uncertainty about the pathway that emissions actually will take. This *scenario uncertainty* also would need to be accounted for if the goal were to predict actual changes.

[7] Note that the requirements of faithfulness and completeness are not completely independent. If a scientist recognizes that a significant source of uncertainty has not been accounted for in a formal analysis, then an uncertainty report consisting of the results of that incomplete analysis will fail to meet the faithfulness requirement as well; it will not reflect his belief about the extent of current uncertainty.

uncertainty reports from individuals, leaving for another discussion the question of how individual and group reports should be related.

Lastly it should be emphasized that, when it comes to informing decision makers, in practice what is important is that uncertainty reports come close enough to meeting the faithfulness and completeness requirements; meeting them perfectly is not usually essential, and in some cases may not be possible or even desirable. For example, if evaluating an additional type of evidence would prolong an assessment beyond when the decision maker needs to make a decision, and if that additional evidence is expected to make at most a very small difference to the conclusions reached in the assessment, then ignoring it may well be a desirable deviation from completeness. What counts as coming close enough to meeting the requirements of faithfulness and completeness? The decision maker should not be led to a make a decision that is substantially different from the decision she would have made with an uncertainty report that did meet the requirements. Since what sort of difference in reported uncertainty would lead a decision maker to a different decision is often unclear, and since the same report may be used for a variety of different decisions by a variety of different decision makers, in general the safest option is to strive for faithfulness and completeness.

3. Two common pitfalls in uncertainty assessment

This section discusses two common pitfalls in uncertainty assessment that can result in uncertainty reports that fail to approximately satisfy either faithfulness or completeness or both. The first involves adopting a one-size-fits-all approach to representing uncertainty, typically in terms of precise probabilities. The second involves ignoring the risk of surprises.[8] Both are prone to result in uncertainty reports that are overconfident, compared to what a more faithful and complete report would indicate.

3.1 Adopting a one-size-fits-all approach to representing uncertainty

In uncertainty assessment, it is sometimes simply assumed from the outset that representations of uncertainty will take a particular form, regardless of the extent of information available. When this happens, most often the assumption is that uncertainty will be represented using precise probabilities – a single-valued probability will be assigned to an outcome or a single probability distribution will be specified over values of a parameter or variable. This appears to have occurred, for example, in the case of the UK Climate Projections 2009 (UKCP09; see also [10]), which aimed to produce high-resolution, probabilistic projections of future climate change for a host of physical quantities, even though it seems clear that current understanding is insufficient to justify assigning precise probabilities for some of these quantities (e.g. for the % change in precipitation on the wettest summer day in London in the 2080s).

There are various reasons why such a one-size-fits-all approach might be adopted. Perhaps, as in the case of UKCP09, it is thought that decision makers desire or require uncertainty reports that provide precise probabilities [11]. Or perhaps the methodologies for estimating uncertainty that are held in highest esteem by the scientist's community call for probabilistic representations, and these approaches serve as a model for the analysis. (Think here of Monte-Carlo-inspired approaches in simulation or standard statistical approaches to quantifying uncertainty in measurement.) Whatever the reason, if uncertainty is represented and reported in terms of precise probabilities, while the scientist conducting the analysis believes that uncertainty is actually 'deeper' than this – e.g. believes that available information only warrants assigning wide interval probabilities or considering an

---

[8] When we speak of the 'risk' of surprise, we mean this in the colloquial sense, not in the narrower sense of having a precisely measurable probability.

5

outcome to be plausible – then the uncertainty report will fail to meet the faithfulness requirement; it will have false precision.

Moreover, this kind of deviation from faithfulness may well make a difference to decision making. One reason is that a precise probabilistic uncertainty report – e.g. "X has probability 0.70" – lends itself to a particular kind of decision-theoretic approach: one that aims to identify optimal policies, such as those that maximize expected utility; without precise probabilities, the decision maker might instead aim for robust policies (see e.g. [12]). In addition, a precise probabilistic report can appear to clearly warrant a particular decision, when a more faithful but less precise report does not. Consider a decision maker who plans to implement a particular policy only if the probability of X exceeds 0.65. An expert report is published which concludes that "X has probability 0.70", but a more faithful report of the estimated uncertainty would have been "X has probability 0.6 – 0.8". Taking the published uncertainty report at face value, the decision maker may proceed to implement the policy, whereas she might not have done so if she were given the more faithful report.

3.2 Ignoring the risk of surprise

Complex systems often exhibit behaviours that are not recognized as serious possibilities by the analytical and heuristic methods that are used to study those systems: they are behaviours that either are deemed exceedingly unlikely by those methods, because the processes that produce them have not been included in the analysis, or else are completely unforeseen by those methods. These behaviours are sometimes referred to as *surprises*. They can be beneficial, adverse or both; of particular interest here are significant adverse surprises, such as serious negative consequences of climate change [13].

There are two main reasons why an agent might encounter surprises. He might have used inadequate methods for exploring the implications of recognized gaps in his knowledge – so-called 'known unknowns'.[9] Recall the modeller in Section 2 who recognizes that there is uncertainty about the values that should be assigned to parameters in his model but who estimates uncertainty about future temperature change via an ensemble study that only varies the initial conditions. The actual temperature change might turn out to be a surprise – it is not among the range of possibilities predicted in his ensemble study – but perhaps it would have been predicted if he'd conducted a more thorough ensemble study in which he also varied parameter values in accordance with their estimated uncertainties. In other cases, surprises stem from unrecognized gaps in an agent's knowledge – so-called 'unknown unknowns'. Here, there are factors shaping the system's behaviour that the agent did not recognize as even potentially relevant – indeed, she may not even be aware of their existence; even if all of the recognized gaps in her knowledge had been filled in, these behaviours would have remained unforeseen. They might be called *genuine surprises*.

In uncertainty assessments, the possibility of surprise is often acknowledged, but not factored into the analysis [14]; the methodology used in effect assumes that there is no risk of surprise. For example, the Reliability Ensemble Averaging (REA) methodology (see e.g. [15], [16]), which delivers probabilistic estimates of uncertainty about future climate change, assigns zero probability to a given interval of change if no model used in the study predicts a change in that interval; the risk of surprise is in effect assumed to be zero. If the risk of surprise is clearly significant, then an uncertainty analysis that ignores it will fail to meet the completeness requirement.

---

[9] As understood here, known unknowns are factors that an agent recognizes as potentially relevant to the question he is addressing but that he has significantly limited knowledge of (e.g. of their presence/absence in the case at hand, or of their strength, or of their precise roles, etc.). As this suggests, the risk of surprise can be different for different agents, or for the same agent at different times, since they/he may possess different information or use different methods of analysis.

Moreover, it may well turn out that different decisions are made than would have been made with a more complete uncertainty report. For example, if a decision maker is made aware that the probabilities generated in a formal analysis are themselves significantly uncertain, or that outcomes other than those predicted are plausible, she might choose policies that protect against a broader range of outcomes. For instance, if her goal is to protect against outcomes that have >0.05 probability of occurrence, she might choose policies that protect against outcomes that, according to the probability distribution provided, have >0.01 probability of occurrence, since she appreciates that outcomes in the tails may be more likely than the distribution would suggest.

4. Interlude: Gauging the risk of genuine surprise

At this point, it might be objected that genuine surprises simply cannot be accounted for in uncertainty analysis, because they stem from unknown unknowns which, by definition, are things that agents do not have any information about. But this is not quite right: it is true that we cannot specify what the unknown unknowns are, else they wouldn't be unknown unknowns; but it doesn't follow that it is impossible to make reasonable judgments about the relative risk of there existing some or other unknown unknown that results in a surprising outcome or behaviour, either when investigating a specific aspect of the system or in the course of investigating many different aspects. On the contrary, there seem to be some clear 'risk factors' for genuine surprise – conditions that, all else equal, tend to increase the chance of an agent's encountering a genuine surprise.[10] If the agent can recognize when those risk factors are present, she might conclude that the risk of genuine surprise is not small and then try to indicate this in her uncertainty report.

4.1 Five risk factors for genuine surprise

What are these risk factors for genuine surprise? Five factors are identified here, though this is not intended as an exhaustive list. These factors relate to: the nature of the system under study, what an agent knows about that system, what an agent thinks he knows about that system, whether the system has shown genuine surprises in the past and whether the system is being subjected to novel conditions.

System Complexity. All else equal, the risk of genuine surprise is higher when the system under study is complicated and nonlinear, and hence complex.[11] This is both because there are more relevant factors and interactions shaping the system's behaviour – and hence more opportunities for unrecognized knowledge gaps – and because even small gaps can give rise to large errors in predictions of the system's behaviour, due to nonlinearities. Likewise, small interventions on the system may be amplified via complicated causal routes – some of which the agent is unaware of – to produce unexpectedly large effects on system behaviour.

Limited Knowledge. All else equal, the risk of genuine surprise is higher to the extent that an agent has limited knowledge of a system – of its past behaviour and the processes that underlie that behaviour. The less the agent knows about a system, and especially the less he knows about the part or aspect of the system that particularly interests him, the more likely it is that he is completely unaware of some of the factors and interactions that will influence what happens.

Overconfidence. All else equal, the risk of genuine surprise is higher to the extent that an agent is overconfident about the extent of his knowledge of the system. That is, the risk of surprise is higher

---

[10] What we have in mind seems similar to what Schneider [13] calls "imaginable conditions for surprise".
[11] While there are different views of what makes a system complex, we assume that being complicated (i.e. involving many interacting parts and processes) and nonlinear is sufficient for being complex.

to the extent that what the agent thinks he knows about how the system works exceeds what he actually knows. In the extreme, he may think that there are almost certainly no unknown unknowns that are relevant to the question that interests him, when in fact there are many.

Past instances of genuine surprise. All else equal, if there have been past instances of genuine surprise when investigating the system, this is a risk factor for genuine surprise going forward. This is direct evidence that the system is capable of presenting the agent with genuine surprises.

Novel conditions. Finally, all else equal, the risk of genuine surprise is higher to the extent that the system under study is being subjected to boundary conditions unlike those in which it was previously studied. In this situation, unrecognized gaps in an agent's knowledge that did not make much difference previously may make a more substantial difference. For example, models of the system that gave reasonable predictions in the past may break down, because – unbeknownst to the agent – they omit feedbacks that have a much stronger impact under the new boundary conditions than under those previously studied. Also, if boundary conditions are changing substantially over time, then the more rapidly they are changing the greater the risk of genuine surprise, insofar as rapid change is more likely to bring about imbalances in the system that exceed the limits of what restoring feedbacks can achieve.

4.2 Can we recognize when these risk factors are present?

From the point of view of uncertainty assessment, what matters is whether an agent can recognize when risk factors like those just identified are present and then form some reasonable conclusion about the risk of surprise when it comes to questions of interest. It is argued here that this is sometimes possible, illustrating with the example of questions about long-term climate change.

System complexity. While the complexity of a system is difficult to quantify, an agent can have substantial evidence that system behaviours of interest are controlled by a large set of processes that interact with one another in complicated and nonlinear ways. This is clearly the case when it comes to earth's climate system, as revealed not only through the study of particular processes and mechanisms but also by the occurrence of abrupt changes in climate in the past (see also [14]).

Limited knowledge. An agent also can have evidence that her knowledge of a system is more or less extensive. If she can make precise predictions of a wide range of system behaviours and can explain those behaviours in a coherent and detailed way, this suggests that her knowledge of the workings of the system is rather extensive. On the other hand, sometimes there is very little that an agent can predict or explain about a system; this is an indication that she has rather limited knowledge of the workings of the system. In the case of earth's climate system, today's scientists seem to be in an intermediate position: they have solid explanations for many climate phenomena and have models that can simulate a variety of salient behaviours, but there are also climate system processes that are not so well understood, such as some cloud feedback and carbon cycle processes, some system behaviours that remain unexplained, and numerous features of the system that cannot yet be simulated/predicted (see also [5]).[12]

Overconfidence. In addition, an agent can have evidence that she is overconfident when it comes to her knowledge of a system, including the knowledge that she considers relevant to addressing a

---

[12] Note that predictive limitations here stem not only from limited knowledge but also from limited computing power; given limited computing power, the climate system must be simulated at a coarser spatiotemporal resolution than desired. Note also, however, that greater computing power need not mean reduced uncertainty, since additional computing power might be directed to more thorough exploration of known unknowns (e.g. via richer models and more comprehensive ensemble studies), which in turn might result in a broader range of simulated responses. In this case, the risk of surprise due to inadequately explored known unknowns might be reduced, even as the model-estimated uncertainty has increased.

particular question of interest. For instance, if it frequently happens that an agent is confident that he's correctly predicted how the system will behave, but his prediction fails, then this is a sign that he is overconfident in his knowledge. (In other words, being surprised a lot in the past, and yet not adjusting one's confidence, is a sign of overconfidence.[13]) In many cases, however, it may be difficult for an agent to tell whether she is overconfident in the knowledge that she considers relevant to a question of interest. This seems to be the case, for example, when it comes to medium and long-term climate prediction; in part because of the long lead-times, today's climate modelers have had few opportunities to see whether their projections are accurate/calibrated.

Past instances of genuine surprise. An agent also can have good reason to believe that a system has displayed genuinely surprising behaviour in the past. This can happen, for instance, if after investigating a surprising phenomenon, an agent comes to explain it in terms of processes or interactions that she previously didn't know existed. An example would be the surprising phenomenon of the Antarctic ozone hole, which was later explained in terms of a previously-unrecognized interaction of chemical (CFCs), meteorological (clouds) and physical (solar) factors (see also [17]). Other times, an agent may know that a system has surprised us periodically in the past, but he may be unsure how many of those surprises resulted from inadequate estimation of uncertainty associated with known unknowns and how many were genuine surprises stemming from unknown unknowns.

Novel conditions. Finally, an agent sometimes can be well aware that the boundary conditions to which a system is subjected are undergoing rapid and substantial change. The climate system is a prime example: there is good evidence that, over the last 150 years, the climate system has been subjected to significant and rapid increases in atmospheric greenhouse gas concentrations, which in turn bring a radiative forcing to the system [1]. As this forcing is increasing over time, so is the risk of genuine surprise in climate system behaviour.

The foregoing suggests that agents sometimes can have good evidence that multiple risk factors for genuine surprise are present. While agents generally will not be in a position to quantify in a precise way the risk of genuine surprise, they may be justified in concluding that, when it comes to questions of interest, the risk is not small and indeed is increasing with time. This seems to be the case with the climate system when it comes to many questions that interest scientists and decision makers, such as questions about the extent of global and regional climate change that would result under different greenhouse gas emission scenarios.[14]

5. Improved uncertainty assessment

If the pitfalls identified above are to be avoided, what should be done instead? How can agents come closer to meeting the requirements of faithfulness and completeness and, more generally, to engaging in good practice in uncertainty assessment? This section outlines some steps and strategies for improved uncertainty assessment in support of decision making.

5.1 Levels of precision, justification and consistency

---

[13] In fact, Morgan ([9], p.7178) notes that a standard measure of overconfidence is the so-called 'surprise index', which is a measure of how often the true value for a quantity lies outside an assessor's 98% confidence interval. (The test focuses on quantities whose true values are known.)

[14] In other situations, we might be able to argue that the risk of genuine surprise is negligible. Suppose I am prone to mild headaches, which taking an aspirin has tended to help relieve in the past. If I now have a mild headache in the same spot as usual (but am otherwise feeling well), and I take an aspirin, it seems reasonable to think that the risk of a genuine surprise when it comes to the outcome (i.e. the risk of something dramatically different occurring due to a factor that I didn't recognize as even potentially relevant) is negligible.

9

The problem with adopting a one-size-fits-all approach to representing uncertainty is that it can easily lead to reports that fail to approximately satisfy the faithfulness requirement, that is, to reports that are significantly misleading about what an agent believes the extent of current uncertainty to be. A better option is for an agent to aim to report uncertainty at a level of precision that matches his belief about the extent to which there is uncertainty. Depending on the matter at hand, this might be in terms of precise probabilities, imprecise probabilities, ranges of plausible values, etc.

It might be thought that representing uncertainty at an appropriate level of precision is simply a matter of introspection: an agent simply "asks herself" how much uncertainty she believes there to be and tries to give an honest description. But as noted above, a request for an uncertainty report is, in ordinary circumstances, a request for a report that is not only faithful but also complete, i.e. that takes account of all significant sources of uncertainty and all available evidence. To come close to meeting the completeness requirement, an agent will typically need to explicitly review, evaluate and synthesize available information, identify the major sources of uncertainty and consider how they interact, and so on. It is thus no surprise that uncertainty assessment as it is actually practiced typically involves activities like these, not mere introspection.

In the end, when it comes to choosing a level of precision at which to report uncertainty – and indeed to formulating the report itself – there are additional steps that can be taken as a sort of "quality control" on the process. For instance, Risbey and Kandlikar [2] outline an approach to choosing a level of precision in which agents are asked to justify their choice. An agent making a preliminary choice to use a full probability density function to represent uncertainty about the future value of a climate variable should then try to justify the 5% and 95% bounds of the distribution as well as its shape (e.g. why this shape rather than an alternative?); if reasons cannot be articulated, the agent should consider moving to a lower level of precision.

Consistency checks are another helpful quality control mechanism. These involve asking questions that can reveal that an uncertainty report is unfaithful. For example, suppose an agent has arrived at an uncertainty estimate that gives a range of plausible values for a climate variable of interest and declares values outside the range to be implausible. The agent might then pause and consider: Suppose the true value turned out to be 10% larger than (or 15% smaller than) the highest (lowest) value in my plausible range; can I imagine how that might happen? If a reasonable explanation can readily be given, the bounds of the range need to be revised [2,9]. Or alternatively the agent might consider: How surprised would I feel if the true value turned out to be 10% larger than (or 15% smaller than) the highest (lowest) value in my plausible range? If the answer is "not very", then again this is a sign that the bounds of the range need revision.

5.2 Accounting for the risk of surprise

The problem with ignoring the risk of surprise is that sometimes it is apparent that this risk is not small; uncertainty reports produced are then incomplete and can be significantly misleading about the current state of knowledge. A better option is to try to take account of the risk of surprise when reporting uncertainty to decision makers. This includes both the risk of surprise due to inadequately probed known unknowns and the risk due to unknown unknowns.

In some contexts, these risks can be learned about empirically. For example, meteorologists can learn about the risk of surprise in near-term probabilistic weather forecasts by examining the performance of the forecasting system over a large set of trials in the recent past. In other cases, including when it comes to projecting long-term climate change with climate models, there is no extensive past track-record to learn from, and it becomes necessary to rely more heavily on scientific understanding, reasoning and expert judgment. For instance, agents can try to identify ways in which the formal methods they've used to probe known unknowns (e.g. ensemble methods) are incomplete or otherwise limited, and then consider whether investigating the implications of those known unknowns more thoroughly would be expected to significantly expand the range of plausible outcomes (see also [13]). Similarly, the analysis of section 4 suggests an approach for gauging the risk of genuine surprises, i.e. those resulting from unknown unknowns: agents can consider whether

multiple risk factors for genuine surprise are present and, reflecting on this, perhaps reach some conclusion about whether the risk of genuine surprise is negligible, non-negligible but small, substantial, etc.

If the risk of surprise is deemed significant, the challenge is then to reflect this in one's uncertainty report.[15] Rarely if ever will an agent be justified in assigning a precise probability to the risk of surprise. This in turn suggests that an appropriate level of precision for the uncertainty report will also be less precise: expressed in terms of imprecise probabilities, plausible ranges, etc. If formal methods of probing known unknowns have delivered a preliminary uncertainty estimate in precise probabilistic form, then factoring in the risk of surprise will usually mean moving to a coarser level of precision; if the preliminary estimate was already expressed in terms of imprecise probabilities, the range of probability may need to be expanded.[16]

Both sorts of adjustment to estimates obtained via formal methods can be seen in IPCC reports. For example, in the Fourth Assessment Report [19], IPCC experts reviewed numerous formal modelling studies that provided 5-95% probability bounds for future global mean surface temperature change under different emission scenarios. Yet the experts ultimately reported their uncertainty in terms of temperature ranges that, while significantly wider than many of the original 5-95% ranges, were only deemed 'likely', i.e. having imprecise probability >0.66 [20]. Similarly, at various points in the Fifth Assessment Report [21], conclusions that formal analyses deemed 'very likely' (i.e. having >0.9 probability) were downgraded to merely 'likely' (>0.66 probability) (see e.g. [22]). Such adjustments were intended to account for sources of uncertainty not sufficiently addressed in the formal analyses. It was often unclear, however, whether these sources of uncertainty included unknown unknowns or just known unknowns that had been insufficiently probed. Such ambiguity reinforces the point that it is desirable for uncertainty reports to be accompanied by a "traceable account" [3] of how they were produced. Arguably, an adequate traceable account in this context would also explain why conclusions were only downgraded to 'likely' rather than, say, 'more likely than not' (i.e. >0.50 probability). This would help to reveal the reasoning behind the scientists' evaluation of the risk of surprise in the particular case at hand.

It is worth noting that better accounting for the risk of surprise in uncertainty assessment is at the same time a means to <u>reducing</u> that risk, relative to what it otherwise would have been. This is because, insofar as the uncertainty reports that are produced attempt to account for the risk, they will be less overconfident. When it comes to projected changes in climate, this also can reduce the risk of surprise 'downstream', when the projections are used by other scientists investigating the impacts of climate change on humans and the environment [23].[17] Of course, there are also other ways to reduce the risk of surprise, even the risk of genuine surprise. For example, substantially cutting greenhouse gas emissions will reduce forcing of the system, mitigating a risk factor for genuine surprise [14,24,25].

6. Recommendations for commissioning bodies

Sometimes failures of faithfulness and completeness ultimately stem not from choices made by scientists during the assessment process but rather from choices made by governmental and industrial

---

[15] What counts as significant depends on the expected uses of the uncertainty report; see Section 2.

[16] An alternative but related approach advocated by Smith [18] involves accompanying model-based estimates of uncertainty with an explicit estimate of the "probability of a big surprise" – the probability of the actual outcome falling significantly outside the range indicated in the model-based analysis. This is in effect a call for reporting second-order uncertainty.

[17] However, it might increase the assessed risk of adverse outcomes since, as noted above, reflection on the possibility of known and unknown unknowns can broaden the range of outcomes that are considered plausible. In this way, a broader range of temperature or precipitation changes may be recognized as real possibilities. See also [13,24,25].

bodies who commission assessments; these bodies sometimes specify parameters for the assessment that make it difficult or impossible for faithfulness and completeness to be met.

For instance, as noted earlier, sometimes a commissioning body requires that scientists consider only a limited range of evidence during the assessment, e.g. evidence from a particular type of study. This will often lead to a failure of completeness. To avoid this, it is advisable that commissioning bodies *avoid overly constraining the evidence base* that experts can consider. While it is clearly undesirable to let poor quality 'evidence' unduly influence conclusions reached in assessments, results that do bear on the matter under assessment should not be excluded from consideration just because they are not of some ideal or preferred type. A prime example, given earlier, is when assessments of medical products and procedures will consider only evidence from randomized controlled trials.

Likewise, sometimes commissioning bodies require that assessment conclusions come in a particular format – e.g. that probabilities are attached to outcomes or statements – which can easily lead to failures of faithfulness.[18] In this regard, it is better if commissioning bodies *allow for flexibility in the reporting of conclusions*. While precise probabilities may seem desirable from the point of view of decision making, requiring scientists to provide them – even if they judge uncertainty to be deeper than precise probabilities would imply – can be counterproductive if the goal is to make decisions that will achieve desired outcomes.

There also are other steps that commissioning bodies can take to help avoid the pitfalls identified above and to support good practice in uncertainty assessment more generally. One such step is to *ask for a traceable account* of how conclusions were reached, preferably one that is brief and non-technical; the account should be considered adequate only if it (i) includes a justification of the level of precision chosen in the reporting of uncertainty and (ii) indicates how the risk of surprise was accounted for in the analysis.

Second, when possible, it is helpful for commissioning bodies to *inform scientists of thresholds that matter for decisions*. This is not always possible, since sometimes decision making options and approaches evolve in dialogue with information provided by scientific assessments. But to the extent that some decision thresholds are clear, this can be useful information for the scientists conducting the assessment. Suppose, for example, that only outcomes that have greater than 1 in 200 chance of occurring are of interest to the decision makers. In that case, scientists might structure the assessment process such that it aims to sort outcomes into those that clearly have a greater than 1 in 200 chance of occurrence, those that clearly have less than a 1 in 200 chance and those whose chance of occurrence might turn out to span the threshold; attention could then be focused on understanding and carefully characterising the uncertainty associated with the latter. Such information about decision thresholds also can help scientists know what sorts of deviations from perfect faithfulness and completeness are unimportant in the context at hand (i.e., what counts as close enough to meeting the requirements of faithfulness and completeness; see Section 2).

7. Why it matters

In closing, this section considers why striving for faithful and complete uncertainty reports, and more generally for good practice in uncertainty assessment, is important. First and foremost it is important because uncertainty reports that fail to meet these requirements are prone to mislead decision makers

---

[18] Steele [26] notes that reporting conclusions in a standardized manner is often required and argues that this requires scientists to make value judgments, at least implicitly, as they decide how to map their epistemic state to the standardized options.

about the current state of knowledge, which in turn may result in their making worse decisions. (By 'worse' decisions, we mean decisions whose outcomes are less desirable by the decision makers' own standards.) This can happen whether reports are overconfident or underconfident.[19] For climate policy decisions, this may mean that serious harms to humans and the environment occur that otherwise could have been prevented.

Second, it is important because, if significant failures of faithfulness or completeness are later revealed (e.g. if uncertainty reports are revealed to be overly precise, implying that more is known than scientists believe is known), then there is a real risk of a loss of credibility for the scientists offering those reports; this may spill over to a loss of credibility for the science itself, even those parts of the science where the evidence is so strong as to warrant definitive conclusions. In the climate context, such a loss of credibility might delay or prevent mitigation and adaptation activities that decision makers would have pursued if they were confident that scientists were providing an accurate picture of the state of climate knowledge. Once again, this may result in greater harms to humans and the environment.

This potential connection with significant harms adds a moral dimension to the task of uncertainty assessment in the climate context. If by taking a bit more care experts can produce uncertainty reports that are substantially less likely to mislead decision makers about the state of climate knowledge, then it seems that they ought to do so. In fact, Douglas [27,28] has argued recently that failing to take sufficient care in arriving at and reporting conclusions – including conclusions about the extent of uncertainty – can constitute negligence on the part of scientists, and thereby expose them to criticism on moral grounds, at least when the lack of care leads to harms that were reasonably foreseeable; the fact that such harms were not intended by scientists does not, according to Douglas, free them from moral responsibility.

The foregoing discussion called attention to two pitfalls of uncertainty analysis that can result in significantly misleading uncertainty reports: adopting a one-size-fits-all approach to representing uncertainty and ignoring the risk of surprise. It was argued that there are steps that can be taken to do better – steps that can help to ensure that uncertainty is reported at an appropriate level of precision, while taking account of the risk of surprise. This includes not just steps on the part of scientists, but also steps on the part of governmental and industrial bodies who commission uncertainty assessments. It is hoped that better awareness of these pitfalls and of the steps that can be taken to help avoid them will contribute to improved uncertainty assessment and, in turn, to improved decision making.

Competing Interests
We have no competing interests.

Authors' Contributions
Both WP and JR contributed to each stage of the work, including conception, analysis, drafting and revising.

References

---

[19] If uncertainty reports are underconfident, implying that uncertainty is greater than scientists believe it to be, then decisions might be delayed, because it is thought that not enough is known to justify action.

Page 13 of 14

Submitted to Phil. Trans. R. Soc. A - Issue
Proc. R. Soc. B article template

13

1. IPCC. 2013 Summary for Policymakers. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley), pp. 3-29. New York: Cambridge.

2. Risbey, J. and Kandlikar, M. 2007 Expressions of likelihood and confidence in the IPCC uncertainty assessment process. *Climatic Change* **85**(1-2), 19-31.

3. 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 & Zwiers, F.W. 2010 *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.* See https://www.ipcc-wg1.unibe.ch/guidancepaper/ar5_uncertainty-guidance-note.pdf

4. Lichtenstein S., Fischhoff B. & Phillips, L. 1982 Calibration of probabilities: The state of the art to 1980. In *Judgment Under Uncertainty: Heuristics and Biases* (eds. Kahneman, D., Slovic, P. & Tversky, A.)*,* pp. 306-334. New York: Cambridge.

5. Risbey, J. & O'Kane, T. 2011 Sources of knowledge and ignorance in climate research. *Climatic Change* **108** (4), 755-773.

6. Parker, W.S. 2010 Whose Probabilities? Predicting Climate Change with Ensembles of Models. *Philos Sci* **77**(5), 985-997.

7. van der Sluijs, J., van Est, R. & Riphagen, M. 2010 Beyond consensus: reflections from a democratic perspective on the interaction between climate politics and science. *Curr Opin Environ Sustainability* **2**(5–6), 409–415.

8. Rowbottom, D.P. 2013 Group Level Interpretations of Probability: New Directions. *Pac Philos Quart* **94**, 188-203.

9. Morgan, M.G. 2014 Use (and abuse) of expert elicitation in support of decision making for public policy. *PNAS* **111**(20), 7176-7184.

10. Murphy, J. M., Sexton, D. M. H., Jenkins, G. J., Booth, B. B. B., Brown, C. C., Clark, R. T., Collins, M., Harris, G. R., Kendon, E. J., Betts, R. A., Brown, S. J., Humphrey, K. A., McCarthy, M. P., McDonald, R. E., Stephens, A., Wallace, C., Warren, R., Wilby, R. & Wood, R. A. 2009 UK climate projections science report: Climate change projections. Met Office Hadley Centre, Exeter, UK.

11. Heaphy, L.J. 2015 The role of climate modelling in adaptation decision-making: the case of the UK climate projections 2009. *Eur J Phil Sci* **5**(2), 233-257.

12. Lempert, R. J. & Collins, M. 2007 Managing the Risk of Uncertain Threshold Responses: Comparison of Robust, Optimum, and Precautionary Approaches. *Risk Anal* **27**(4), 1009-1026.

13. Schneider, S., Turner, B. & Garriga, H. 1998 Imaginable surprise in global change science. *J Risk Res* **1**(2), 165-185.

14. Schneider, S. 2004 Abrupt non-linear climate change, irreversibility and surprise. *Global Environ Change* **14**(3), 245-258.

15. Giorgi, F. & Mearns, L.O. 2002. Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'Reliability Ensemble Averaging (REA)' method. *J Clim* **15**, 1141–1158.

16. Xu, Y., Gao, X., & Giorgi, F. 2010 Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. Clim Res **41**, 61-81.

17. Solomon, S. 1999 Stratospheric Ozone Depletion: A Review of Concepts and History. *Rev Geophys* **37**(3), 275-316.

18. Smith, L.A. 2012 Predictability and Insight: Contrasting the achievable aims of forecasting in weather-like cases and climate-like cases. Discussion Draft. See http://www.forecasters.org/proceedings12/Smith_Len_ISF_Boston.pdf

19. IPCC. 2007 *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor & Miller, H.L.). New York: Cambridge.

20. Meehl, G.A., T.F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M. Gregory, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver & Zhao, Z.-C. 2007 Global Climate Projections. In [10], pp. 747-845.

21. IPCC. 2013 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley) New York: Cambridge.

22. Bindoff, N.L., P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari & Zhang, X. 2013 Detection and Attribution of Climate Change: from Global to Regional. In [21], pp. 867-952.

23. IPCC. 2014 *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea & White, L.L.) New York: Cambridge.

24. Lewandowsky, S., J. Risbey, M. Smithson, B. Newell & Hunter, J. 2014 Scientific uncertainty and climate change: Part I. Uncertainty and unabated emissions. *Climatic Change* **124**(1-2), 21—37.

25. Lewandowsky, S., J. Risbey, M. Smithson & Newell, B. 2014 Scientific uncertainty and climate change: Part II. Uncertainty and mitigation. *Climatic Change* **124**(1-2), 39-52.

26. Steele, K. 2013 The scientist qua policy advisor makes value judgments. *Phil Sci* **79**(5), 893-904.

27. Douglas, H. 2003 The Moral Responsibilities of Scientists (Tensions between Autonomy and Responsibility). *Am Philos Quart* **40**(1): 59-68.

28. Douglas, H. 2009 *Science, policy, and the value-free ideal*. Pittsburgh: Pittsburgh.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60