Numbers: Gods, Certainty and Science
Scraps from the Philosophers' Banquet

Roger M. Cooke

Oct 4, 2024

Preface: an Interview

1. **Numbers: Gods, Certainty and Science**

2. **Gods**
   2.1 **Agency: The Illiad**
   2.2 **Immediacy: Passez outre**
   2.3 **Dissimulation: Islamic Pluralism**
   2.4 **Midrash: When Trading Fails**

3. **Certainty**
   3.1 **Partial Belief and Messaging Climate Change**
   3.2 **Shallow Uncertainty**
   3.3 **Scientists Sentenced to 6 Years in Prison for "Optimizing" Uncertainty of L'Aquila Earthquake**

4. **Science Under Uncertainty**
   4.1 **Conflicting Measurements Reduce Uncertainty in Climate Sensitivity**
   4.2 **Respectability Looming for Expert Judgment**
   4.3 **The Science of Forecasting: the Special Sauce**
   4.4 **The Iceman Cometh**
   4.5 **Wisdom of Crowds**

5. **The Hard Part: Twining**

References

Preface

This book gathers attempts of a peripetetic philosopher to communicate to a wider public. It is gathered from various blogs and public talks, hyperlinked in the text, based on scientific source listed at the end. It is written to appease my personal daemon; if others find it worthwhile, that is a very welcome co-benefit. Fragments from a recent interview[1] set the stage.

You originally studied Philosophy of Science at Yale – can you tell us a little bit about that, and about how that study influenced the later development of your thinking? For a Philosopher you have done a lot of work on real life problems.
I started in Philosophy of Physics. Not only was that program weak at Yale, but I struggled with the basics. Put a glass of water on a table. There are two invisible forces acting, gravity is pulling the glass down and the table is pushing back up with, miraculously, the exact same magnitude in the opposite direction so that nothing happens. Take away the table and the glass falls, take away gravity and nothing happens. To do the simple physics exercises you have to know the code, and know which questions not to ask. I eventually learned the code, but retained the sense that there were flaws in the story. The real puzzle for me was why "force" worked and "phlogiston" didn't. Why "space" and "time" worked but "ether" didn't. I switched to Philosophy and spent a lot of time on the Greeks, the Scholastics and Enlightenment philosophers, especially Kant and Hegel.

The great philosophers construct a coherent system for understanding everything. In so doing they start with the natural language and progressively re-wire it so that concepts successively acquire new meanings, defined in evolving contexts. You can't understand it piecemeal; you just have to keep going until it all starts making sense. Once you "get it" you can see everything in a new way, like a conversion experience. Most people have at most one conversion experience which they then regard as apodictic. If you study philosophy you go through several….it helps. In retrospect, that's one of the great things I learned in philosophy, that and how to read seemingly unintelligible texts.

Here's an anecdote: a math colleague and I were trying to learn atmospheric dispersion modelling. Atmospheric chemists have their own code, which mathematicians find inscrutable. We started with an elementary text. The colleague would come to something he didn't understand, stop and look for another text to explain. That sequence doesn't converge. I would just keep reading until their code starts to become intelligible.

The great Systematic Philosophies have at their core a theory of knowledge. What knowledge is determines what we can know; 'what we know' and 'how we know' are very tightly coupled. For Plato, knowledge was acquired by direct intuition of a soul sufficiently purged of false beliefs. For the Scholastics, knowledge is Reason applied to Divine Revelation. For Kant, Newton's mechanics and Euclidean geometry enjoyed a level of certainty not attainable by induction from observations: they are necessarily imposed on our perceptions of nature by our knowledge apparatus – or so he thought. He was wrong about that, but he was right that, to turn a philosophical phrase, knowledge and the knowledge of knowledge are inseparable. For Hegel, knowledge is self-consciousness of trans-personal spirit - another story altogether (see chap 5).

Foundations of probability played a big role in your thinking, could you elaborate?
The flaws in classical mechanics began to extrude themselves in the latter 19th century and people felt that language was a big part of the problem. In an effort to separate pure definitions and mathematics from deliverances from experience Heinrich Hertz gave the first axiomatization of classical mechanics in a proto-formal language. He found the notions of

[1] Cooke R.M. (2021) Building on Foundations: an interview with Roger Cooke, in Expert Judgement in Risk and Decision Analysis eds Nane, Hanea, French and Bedford, Springer Nature Switzerland AG, Cham, Switzerland.

force and absolute space-time superfluous and unhelpful. Such formal approaches were cross fertilized with activity at the foundations of mathematics – another story. Ernst Mach invented "semantic analysis" whereby notions must obtain a semantic pedigree tracing them to elementary sensations before they are serviceable to science. It emerged that concepts like phlogiston, force, absolute space-time lacked semantic pedigrees. Propositions assigning them properties are not unknowable, they are meaningless. The power of that insight emerges when you contemplate all the unknowable things people believe. The revolutions of relativity and quantum mechanics drew heavily on semantic de-constructions. Mach himself believed that atoms also lacked a semantic pedigree and that propositions about atoms were therefore meaningless. Atoms, however proved very useful. Indeed where would modern physics be if confined to Mach's semantic strictures?

Philosophy of science emerged as an effort to articulate the scientific method and thereby determine what science is and is not. Is risk analysis science? psychoanalysis? creationism? economics? Terms like leptons and quarks do not have operational meaning in the narrow sense as they are not directly linked to measurements, yet they seem to be ok. What about Freud's id? creationism's intelligent designer? economists' representative consumer? What about randomness? fuzzy membership? degrees of possibility? You see where this is going. The demarcation of science and non-science is closely bound up with the problem of "theoretical terms": articulate a semantics in which terms without direct operational meaning, nonetheless acquire meaning in a given theory. There is a load of active literature on this, which I have tried to boil down to a simple formula (see Cooke, 2004) "The operational meaning of "*degree of possibility*" in the proposition: *"The degree of possibility that the Loch Monster exists is 0.0031416"* is the set of non-tautological propositions not containing "*degree of possibility*" which that proposition implies.

What about "uncertainty", what does it mean? In the natural language it means different things in different context, including ambiguity, ambivalence, confusion, distrust, unpredictability and indecisiveness. Anyone wishing to "represent uncertainty" in a scientific context must do some serious re-wiring. As often happens, a scientific reconstruction of a term in the natural language captures only part of its native meaning. Compare "force" in physics and in the natural language. L.J. Savage's foundation of subjective probability is a superlative example of rational reconstruction in science. He provides axioms describing rational preference with clear operational meaning for the primitive terms (see Chap 3.1). Strong arguments support his axioms – maybe not as strong as arguments for the axioms of Zermelo Frankel set theory[2], but very strong nonetheless. He then proves that the preferences of a rational individual can be represented in terms of a personal probability (aka subjective degree of belief) which is uniquely determined and the utility function which is unique up to a choice of zero and unit. All my students had to learn these proofs, not only to understand uncertainty but also to understand how to extend the purview of science. Others may protest that uncertainty means much more than subjective probability. Duh. However, if you want to quantify, say ambiguity, you must provide operational meaning telling us whether "degrees of ambiguity" are positive, how they are measured, whether they can be added or multiplied, etc. Those properties must be derived from the operational meaning of the primitive terms. At a conference in Paris, a leading light presented his new definition of uncertainty which unbeknownst to him, allowed uncertainty to take negative values. The theologians would love that. There have been many variations on Savage's axioms, just as there have been many variations on Zermelo Frankel set theory, but they all remain variations around a core theory that is suitable for applications. There are also countless "alternative representations" of uncertainty which lack any foundation whatsoever.

How did the idea of a rational consensus emerge – can you describe what it is and why you think it its useful to policy and decision makers?

We come to the theme of extending the purview of science. Traditional philosophy of science pretends that, within the context of justification, science deals only with certainties and reasons deterministically. It isn't so. Society is increasingly confronting decisions with large uncertainties with consequences impacting our survival. We all know the myriad ways in which private interests can and will exploit uncertainty to further their own interests. We must bring 'decision making under uncertainty' within the purview of science. Savage provides necessary but not sufficient conditions for rational decision making under uncertainty. Indeed, ANY subjective probability combined with ANY utility is rational in the sense of

[2] These axioms (including the axiom of choice) are generally accepted as the basis for modern mathematics.

Savage. Rationality in science, whatever that means, is much more restrictive. The challenge is to bring science based restrictions into Savage's model, at least with respect to probability, such that all subjective probabilities are not equal. Utility is another problem. Validation is not hopeless but much less active than the probability component of rational decision (see Neslo and Cooke, 2011).

I first encountered the term rational consensus in a book by Keith Lehrer and Carl Wagner (1981). It is similar to that of De Groot (1974), discussed in Cooke (1991). Participants assign probabilities to events and weights to each other's probabilities, leading to an equilibrium distribution. There's nothing scientific about it IMO, and it is not remotely practical. Experts are over worked and under paid. They're not going to travel long distances to sit together and reach 'dialectical equilibrium' as prerequisite for weighing each other.

However, the term rational consensus stuck in my mind and I sought a more science-based meaning[3]. The idea is that experts construct their rational consensus. They quantify their degrees of belief as subjective probabilities for both the variables of interest and for calibration variables taken from their field. Such quantification typically takes the form of stating a median value (with probability ½ of above and ½ below the true value) and a 90% uncertainty range (with probability 90% of catching the true value.

Experts are scored as statistical hypotheses with respect to statistical accuracy and informativeness (see Chapt 4.2). The theory of strictly proper scoring rules, appropriately generalized, converts their scores into weights. This means that an expert maximizes his/her expected weight by and only by stating his/her true opinion. The combination scheme satisfies necessary (not sufficient) conditions for the scientific method. The necessary conditions are *traceability, neutrality, fairness* and *empirical control*. Traceability means that all steps in the calculation must be open and reproducible. Fairness excludes pre-judging experts. Neutrality corresponds to proper scoring rules. Empirical control is the most consequential: It requires that experts quantify their uncertainty for variables from their field for which the analyst (but not the expert) knows or will know the true values. The frequency with which the true values fall inside or outside the expert's 90% ranges, and above or below the expert's medians is tabulated. As this frequency departs from what we should expect from a statistically accurate expert, the expert's weight goes down; when a threshold is crossed the expert receives weight zero. The possibility of assigning zero weight to statistically very inaccurate experts implements Popper's idea of falsifiability.

Rational consensus means that experts pre-commit to the results of the combination. They needn't adopt the result as their personal probability. However, withdrawing from the rational consensus imposes a proof burden of showing how the necessary conditions were violated or were insufficient. Tony O'Hagan's question 'is rational consensus a subjective probability, if so whose?' gets the simple answer: it is the personal probability of any rational agent whose personal probability agreeing with the rational consensus.

Can you tell us something about the types of risk problems that you were thinking about when you started developing your ideas about expert judgement?

The topology of the problems was defined in the Rassmussen Report (USNRC, 1975) and evolved through three generations as described in (Cooke, 2013) . We have panels of order 10 experts assessing up to 100 uncertain quantities. Discrete events are sometimes assessed, but most variables are effectively continuous. The Rasmussen report did a good job on traceability. Publishing all the expert raw data made visible the very large differences between experts, thereby teeing up of combination and validation. The Rasmussen report selected the distributions used in the report in a rather inscrutable fashion. In the second generation studies, experts' rationales were catalogued and their distributions were combined with equal weighting. The third generation in which I participated added performance measurement, empirical validation, dependence modelling and probabilistic inversion.

[3] This discussion is highly simplfied, a full mathematical discussion is in Cooke (1991), a simplified modern exposition is in (Cooke et al 2021) especially the supplimentary online material.

Finally, if you could organize a dinner party with 3 or 4 'great thinkers' who influenced your development of the classical model, who would you invite and why?

Learning to reason probabilistically will be an event in the cognitive history of Man comparable to the formulation of deterministic reasoning in Aristotle's Logic. The great hero here is Frank P. Ramsey. His "Truth and Probability" (1926) is a bolt of sheer genius. Let's also include John von Neuman (Theory of Games and Economic Behavior, 1944) and Lenard Jimmy Ogashevitz (aka Savage)(The Foundations of Statistics, 1954). But not for dinner - nobody could get along with von Neumann. The most important people at the inception of CM were Louis Goossens, Max Mendel and Simon French. Early adapters from the first hour were Willy Aspinall, Tim Bedford, Jan van Noortwijk, Matthijs Kok, Dmitri Solomatin, Gordon Woo, Tom Mazzuchi and Christian Preyssl. Follow on forces include Dorota Kurowicka, Anca Hanea, Tina Nane, Oswaldo Morales, Jim Hammitt, John Evans, Abby Colson, John Quigley, Justin Eggstaff, Rene van Dorp, Arie Havelaar, and Ben Ale. These would also need to be invited; we will need a Banquet Hall. Then we can also invite all the colleagues who performed the applications, Kim Thompson, Radboud Duintjer Tebbens, Juoni Tuomisto, Nicole van Elst, Daniel Puig, Frank van Overbeek, Xi Quing, Maurits Bakker, Rabin Neslo, Daniel Lewandowski, Sandy Hoffmann, Matt Gerstenberger, Maart Janssen, Augusto Neri, Eric Jager, Ben Goodheaart, Juliana Lopez de la Cruz, Julie Ryan, Maartin Nauta, Marion Whitmann, David Lodge, John Rothlisberger, Arno Willems, Jim Smith, Fred Harper, Steve Hora, Mark Burgman, Elizabeth Beshearse, Raveem Ismail, Vicki Bier, Bernd Kraan, Ben Koch, Daniela Hanea, Christoph Werner, Bis Bhola, Michael Oppenheimer, Jonathan Bamber, Bob Kop, Monika Forys, Michael Tyshenko, Maartin Nauta, Karin Slijkhuis, Kevin Rennert, Richard Newell, Sarah Teck, Ben Halpern, Michael Tyshenko, Elizabeth Beshearse … with apologies to everyone I forgot. Recalling all these people and their contributions is quite humbling. BTW, didn't we have just such a banquet in July 2017?

1. Numbers: Gods, Certainty and Science

Adapted from Oration "A Number of Things" delivered in accepting the position of Professor in Applications of Decision Theory at the Faculty of Technical Mathematics and Informatics at the Delft University of Technology, Delft, November 8, 1995 by Dr. R. M Cooke. Translated from Dutch by the author.

I hail from the United States, a country where the custom of delivering an oration is unknown. From all the no-nonsense Yankee business jargon currently emanating from Dutch university administrations, I inferred that the oration had come to resemble a Medieval morality play, somehow out of step with the times. I am a parvenu Dutchman, and people whom I judge much wiser than myself have convinced me, not without a certain impish pleasure, that I too must give an oration.

But how? A study of the genre reveals that the ideal oration opens with a quote; a quote which surprisingly explains a seemingly nondescript title by linking the Aspirant Professor's field to large themes from, preferably, Dutch history; and this all with a bombastic intellectual swagger which somehow stays entertaining. I shall try to perpetuate this tradition. A nondescript title was easily found (the oration was titled "*A Number of Things")*. Long did I search for the quote. It appears that history's key figures seldom refer to mathematics. Has mathematics had nothing to say to them?

Some hope could be gleaned from Max Weber's *Die Protestantische Ethik und der Geist des Capitalismus.* Weber uncovered a strong link between the origins of capitalism, the industrial revolution and the emergence of Dutch Calvinism. The Calvinist doctrine of predestination, as you know, had the effect of devaluing the most important asset of the Catholic Church, real estate in the Here After. The decision who would go to Heaven and who would not, had already been taken, and the Church could not intercede. Moreover, those who had been elected for salvation could not be identified by any outer or inner property. What is then the point of this short life on Earth? Our only earthly goal must be to nurture hope for an undeserved salvation. In The Netherlands, that translated to earning as much money as possible without enjoying it. There was no alternative but to apply the unspent gain to garner yet more gain, and capitalism was born, according to Weber. The hallmark of the spirit of capitalism, says Weber, is that everything, but then really *everything,* should be calculated in terms of capital. As spokesman par excellence of this new spirit, Weber cites an erstwhile compatriot. *"The good paymaster"* says Benjamin Franklin

"is lord of another man' s purse. He that is known to pay punctually and exactly to the time he promises, may at any time, and on any occasion, raise all the money his friends can spare"[4]

Perhaps Prince William of Orange recognized in this spirit of capitalism the possibility of an alliance between Calvinist ministers and the Dutch sea pirates, from which the State of The Netherlands

[4] Weber, M, The Protestant Ethic and the Spirit of Capitalism, Scribner's Sons, New York, 1958, p. 49

eventually emerged.

If such calculation does lie at the basis of the Dutch nation state and the creation of modern credit worthy man, then the large themes emerge; religion, the origin of nations, and numbers. What hidden relations bind these concepts? How is the Earth divided into "We's" and "They's"? Why are there nations and Gods, why so many, and for how long? Though Franklin's quote evokes all these questions, in no way does it cover the activities of the Chair of Applications of Decision Theory. Deeper must we dig.

It appears from the first European national anthem[5] that the founding of The Netherlands is intimately bound up with the gift of God to David of "a kingdom in Israel, most great".

How was that exactly? The founding of Israel is symbolized in the founding of the Temple of King David in Jerusalem. The story is told in the Bible, First Chronicles, chapters 20-22. In his last battle, David defeated several Ammonite cities. He led the inhabitants out and "cut them with saws and with harrows of iron and with axes", in accordance with the wishes of the Lord. Shortly thereafter, however, he listened to Satan and ordered the Israelites to be *counted* . The wrath of the Lord was immediate. David was given a choice, "either three years famine, or three months to be destroyed before thy foes, ... or else three days the sword of the Lord". David chose the latter, and seventy thousand Israelites were laid low by God before David repented (he was allowed to count the dead). The angel of the Lord showed David the spot where he should build an alter to the Lord, and on that spot the Temple of Jerusalem was built.

The roots are laid bare. Imagine, ladies and gentlemen; the pictures are familiar from the daily news. Naked children torn from their mothers' breasts, children scream, mothers plead; but the Lord is implacable and the saw teeth of the Lord chew on. For indeed, those children would have grown up worshipping a different God. David need show no remorse for this ethnic cleansing. He is unfaithful to the Lord only when he counts the number of his own people. David counted the Israelites because, like any commander, he wanted to know his military strength, but he should have known that his strength came solely from the Lord. The Lord would deliver him if he put his faith in the Lord. Trying to take his fate in his own hands was high blasphemy. Sawing the children of the enemy to pieces did not incur the Lord's displeasure. At a technical university we count, calculate and measure to gain control over our fate. In my field of risk analysis we attempt daily to frustrate the 'acts of God'. Is that too high blasphemy?

The founders of nations renounce existing earthly law, and appeal to incontrovertible supernatural authority. That's the way it has always been, and that's the way it is today. How does science ultimately relate to the fruits of such labor? This is the old question of the relation between reason and authority, between science and faith. During the Enlightenment the ethical basis of modern constitutional democracy was negotiated by, among others, Immanuel Kant. Kant's answer came down to an armed truce between reason and faith. Each was assigned its own territory and instructed not to pester the other. Can this compromise hold its own in the face of the continual re-allocation of the earth? The problem is that the various incontrovertible authorities cannot leave each other alone, and if reason is kept out, then only the saws, iron harrows and axes remain.

The question of the relation of reason and authority receives a much more radical answer in a casual aside

[5] The Wilhelmus van Nassouwe; see Schama, S. (1987} The Embarrassment of Riches: an Interpretation of Dutch Culture in the Golden Age, Fontana Press, London, p. 103. The Dutch of ten emphasized the analogy between the Israelites and their own quest for nationhood, as reflected in the eigth stanza of the Dutch National Anthem, the first European national anthem.

of the Danish physicist Niels Bohr. His comment also perfectly describes what we in Applied Decision Theory try to do.

One day, Bohr visited the Danish Parliament as guest of an eminent politician. A heated discussion was under way, and his host remarked "...this is certainly quite different from the discussions at your institute, is it not Professor Bohr?" Bohr answered that discussions at his institute also became quite heated. He paused for a moment and added "...but there is one difference, at our institute we try to agree"[6].

"Is that all?" I hear you ask. Yes, that is all. Gods do not try to agree. Allah and Jehovah will never agree which incontrovertible authority is the true one. Politicians make compromises, that is, they find power equilibria. Scientists, on the other hand, agree. If the founding of nations is bound up with appeals to incontrovertible supernatural authority, then science is building a sort of anti-nation. Science creates a "we" which is not based on mutual recognition via a commonly recognized authority, but it is based on something else. And what is that ladies and gentlemen? Numbers. Numbers are the things on which homo sapiens can agree. We in decision theory try to replace discussions about power and authority with discussions about numbers.

Applications of Decision Theory

Let me explain. When my daughter studied at the Royal Conservatory of Ballet, we once took a vacation in the mountains. We chanced upon a deep ravine over which a large tree had fallen. Dear daughter jumps on the tree and starts across. "If you fall off you will never dance again" advise I. "But I won't fall off" she answers indignantly. I could have appealed to my parental authority, but then I would always remain the father who forbade the tree. Instead, I applied decision theory. "Okay, go ahead if you must, but first estimate the chance that you will fall, is it one in a hundred, one in five hundred? tell me." Daughter reflects for a moment and climbs off the tree.

Once we start counting people, we don't stop. I have here a graph showing the world population from 10,000 years ago up to the present[7]. The graph begins with a population of 10 million in 8,000 BC and creeps slowly upward until the year 1650, then suddenly it shoots up. Before 1650 the world population grew at the rate of 50% per thousand years, every 1000 years it increased by 50%. After 1650 it increases at the rate of 2000% per 1000 years.

[6] Personlalanecdote of Prof. H.B.G. Casimir

[7] Hauser, P.M. (1975) "World population problems" Headline Series Foreign Policy Association no. 174.

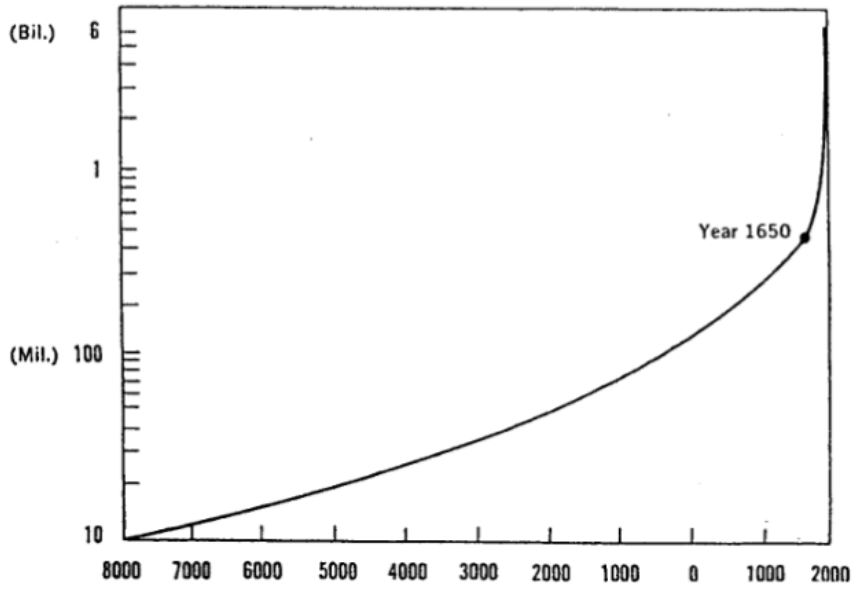

Figure 1 World population from 8000 BC

What explains this kink around 1650? Dutch Calvinism perhaps? Alas I must disappoint you. On a scale of 10,000 years there have been hundreds of Hollands, hundreds of Calvins, and hundreds of people who returned from the Dead. Yet there is only one kink. We are dealing here with the anni mirabiles between the publication of Copernicus' De Revolutionibus Orbium Caelestium in 1543 and the Philosophiae Naturalis Principia Mathematica of Newton in 1687. These are the years in which modern science and the industrial revolution were born.

What is going on? During the anni mirabiles a unique event occurred in the West. Everywhere there was technology, the fabrication of tools, and many cultures possessed some form of science. At this time in the west, however, the two came together. The marriage between science and technology meant in the first place that scientists acquired better instruments with which they could discover natural laws. Knowledge of these laws enabled them to make more accurate instruments, with which they could discover still more laws, make better measurements, etc. Better instruments served not only for better measurements. They also provided better navigation, better methods of production, better agriculture; more people could be fed with less labor. There was more free time for still more improvements, and thus 2000% per 1000 years.

The 'scientization' of technology is an event which is visible on a time scale of 10,000 years. The activities of Applied Decision theory are not visible on this scale, but they are visible on a scale of 30 kilometers.

The figure below shows the lateral spread of a plume of airborne radioactive material after a hypothetical accident at a nuclear power station under stable atmospheric conditions in northern Europe. Despite intensive efforts of large research laboratories like Kernforschungzentrum Karlsruhe[8] and the National Radiological Protection Board[9] ; the prediction of such a plume spread still requires a raft of uncertain

[8] Fischer, F. Ehrhardt, J. and Hasemann, I. (1990) Uncertainty and Sensitivity Analyses of the Complete Program System UFOMOD and of Selected Submodels.

[9] Crick,, J.J., Hof er, E. Jones, J.A. and Haywood, S.M. (1988) Uncertainty analysis ofthe Foodchain and Atmospheric Dispersion Modules of MARC. National RadiologicalProtectionBoard,report184, Kernforschungzentrum Karlsruhe, Report 4627

parameters.

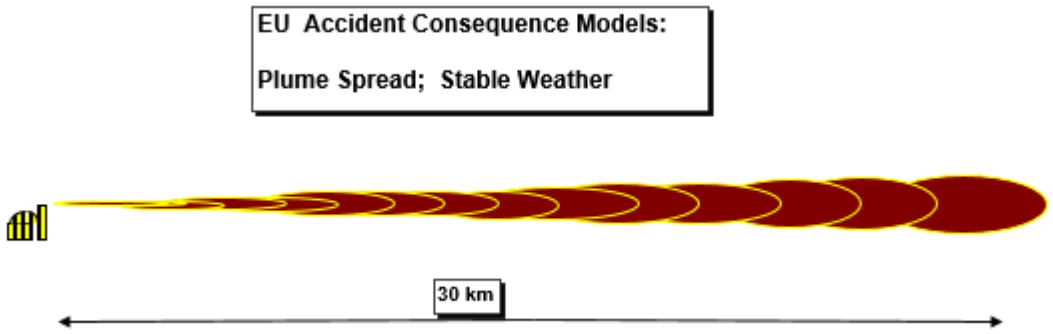

Figure 2 Lateral plume spread under stable atmospheric conditions

In the 1980's the research labs performed various 'uncertainty analyses' of consequence models. The uncertainty in the input parameters was quantified and propagated through the models. The resulting uncertainty in model predictions can be summarized in 90% uncertainty bands. The next figure illustrates the 90% uncertainty bands for lateral plume spread under stable conditions. According to these analyses, we may be 90% certain that in a real accident under these conditions, the lateral plume spread will lie between the upper and the lower plumes.

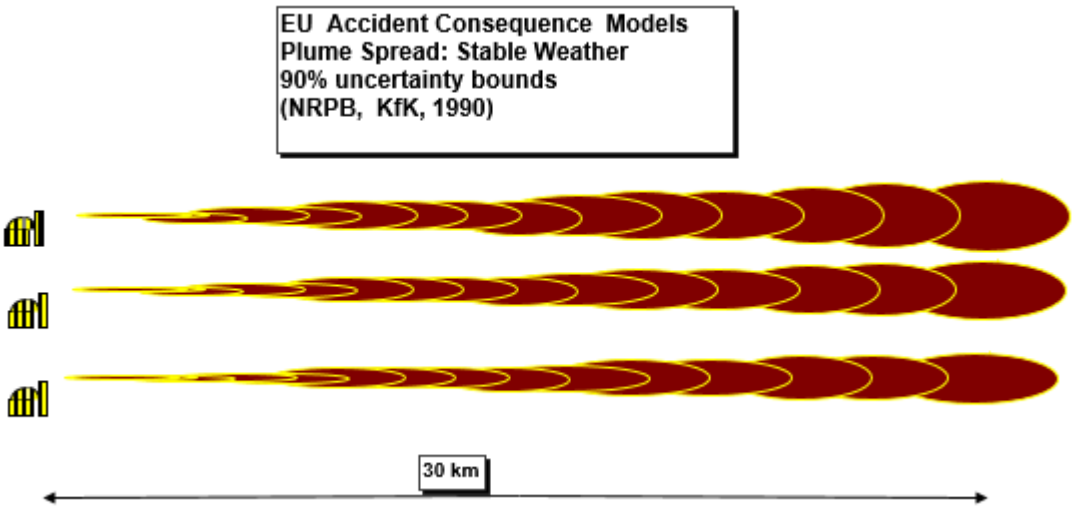

Figure 3 90% uncertainty bands for lateral plume spread under stable atmospheric conditions

It will be noted that these uncertainty bands are rather narrow. These were obtained with the 1960's BOGSAT method (Bunch of Girls/Guys Sitting Around a Table (see chapter 4.4). The scientists are quite certain of the degree to which they can predict the plume spread. Is this degree of certainty justified? Such questions can easily degenerate into discussions of power and authority. In 1990 a joint research program was initiated between the European Union and the American Nuclear Regulatory Commission (USNRC). The goal was to redefine the state of the art regarding the uncertainty analysis of large scale consequence models.

In the course of this project, uncertainties for input and output variables for European and American models were agreed upon. A large number of European research labs participated, and overall coordination of the

European effort rested with the Safety Science group in Delft. The chair of Applications of Decision Theory provided mathematical support.

The analysis of uncertainties in large risk models involves many interesting mathematical and philosophical questions. One of these lends itself for illustration here. By way of introduction, I show our results for the uncertainty in lateral plume spread under stable conditions[10].

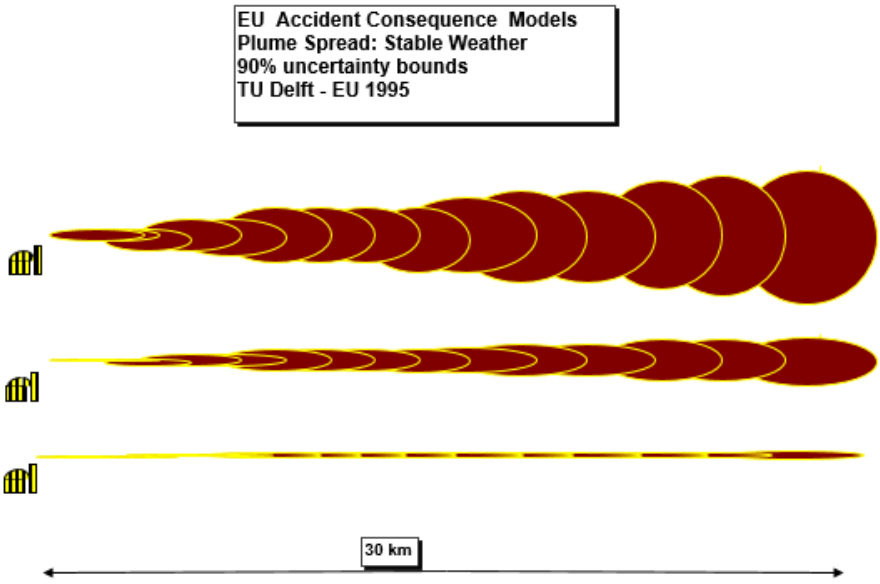

Figure 4 90% uncertainty bands for lateral plume spread under stable conditions;TU Delft method

Comparing the previous two figures, it is evident that a new picture of the uncertainties has emerged. If you reflect that the seriousness of an accident is determined in large measure by the degree to which the plume does not spread, then you can imagine the consequences of this new picture for emergency planning.

This new picture emerged the way new science often emerges, with new instruments. The Dutch invention of the telescope in 1608 opened a world of scientific observations. People first validate the telescope by training it at things afar which we already know, then we apply it to things like the planet Venus which we don't know. Here, the new instrument is the scientists' own assessment of his/her subjective uncertainty. Scientists provide a 'best guess' having an even chance of being above or below the true unknown value, and they also provide a 90% confidence band, an interval having a 9-out-of-10 chance of catching the true value. Each participating scientist is a new instrument and like all new instruments it must first be validated against things we already know. Telescopes are not all created equal and the same holds for our new uncertainty assessing instruments. We applied the the uncertainties represented in Figure 3 to 36 real tracer experiments in which plume spreads under the relevant atmospheric conditions were measured. In 20 of the 36 experiments, the measured values fell outside their respective 90% confidence bands. If each 90% confidence bands really had an independent 90% chance of catching the true value, then the probability that 20 or more observations fall outside these 36 bands would vanishingly small. Such results motivated the performance based combination of expert judgments developed at the TU Delft[11]. Experts are weighed according to their performance as uncertainty assessors: How statistically accurate and how informative are their assessments on questions from their field to

[10] Cooke, R.M. (1994) "Uncertainty in dispersion and deposition in accident consequencemodelingassessedwithperformance-basedexpertjudgment" Reliability Engineering and System Safety no. 45 35-46.
Cooke, R.M., Goossens, L.J.H., and Kraan B.C.P. (1995) Methods for CEC/USNRC AccidentConsequence UncertaintyAnalysisofdispersionand Deposition. EUR-15856 EN.
[11] Cooke, R.M. (1991) Experts in Uncertainty,Oxford University Press.

which true values are or become known? Determining these weights is some times a humbling, not to say therapeutic, experience.

Much has been learned about these new uncertainty assessing instruments; what makes a good instrument? how good are they? what contribution do they make? This is unraveled in following chapters, here a spoiler alert: There are real differences in performance of our uncertainty assessing instruments, they should be combined judiciously to get the best results. Their main contribution is in the quantification of risk and supporting risk-based decision making. Examples abound in nuclear safety, commercial aviation, public health, infrastructure, pandemics, invasive species, natural hazards and of course the greatest risk of all, climate change[12].

One last remark on this example before concluding: Colleagues, especially colleagues in the social sciences often wonder how world-renowned experts can be scored on performance as if they were school children. People without a background in the empirical sciences are surprised to hear that the experts actually enjoy this. The overwhelming majority of experts appreciate any attempt to replace discussions of power and authority with discussions of numbers, even if it concerns their own power and authority. They would all feel very much at home in Bohr's institute.

Conclusion

Our culture still needs symbols of incontrovertible authority. A striking example of this is closer than you may realize. During a recent 'professors dinner' I learned that when a professor dons his/her cap, then he/she exercises his/her official function and cannot be contradicted. By delivering this oration with my cap, I am an accomplice in this symbolism. Is that entirely consistent with the aim of replacing discussions of authority with discussions of numbers? After extended internal debate, I concluded that I could consistently wear this cap[13], for the following reason. Challenging symbols of incontrovertible authority does not reduce the need for such symbols. If this need emanates from fear, then such a challenge only amplifies the fear and thus intensifies the need. What is the antidote for fear? Socrates prescribed irony. After all, what is more ironic than a scientist with a cap posing as incontrovertible authority? Socrates made a distinction between irony and hypocrisy ...by drinking the hemlock[14]. In the long run, however, there is only one cure for fear, and that is knowledge.

But how long is the long run? I return to the picture of the world population from 8000 BC. I have here the same picture, but now the time axis is extended out to 8000 AD. Mathematicians like to extrapolate; how should we extrapolate the world population line out to the year 8000 AD.

[12] See Cooke, Roger M., Marti, Deniz and Mazzuchi, Thomas A., (2021), Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say?" https://doi.org/10.1016/j.ijforecast.2020.06.007 for a recent overview.

[13] The cap pictured at the beginning is actually out of compliance – it is a Portuguese bull fighter's cap.

[14]The Dutch expression"ergens gif op innemen" (to take poison on it) means roughly "to bet your life on it". The play on words in this context is untranslatable. Socrates' irony was in deadly earnest. After being found guilty of corrupting the youth by teaching them to question authority, he surprised his followers by refusing escap and drinking the Hemlock poison.

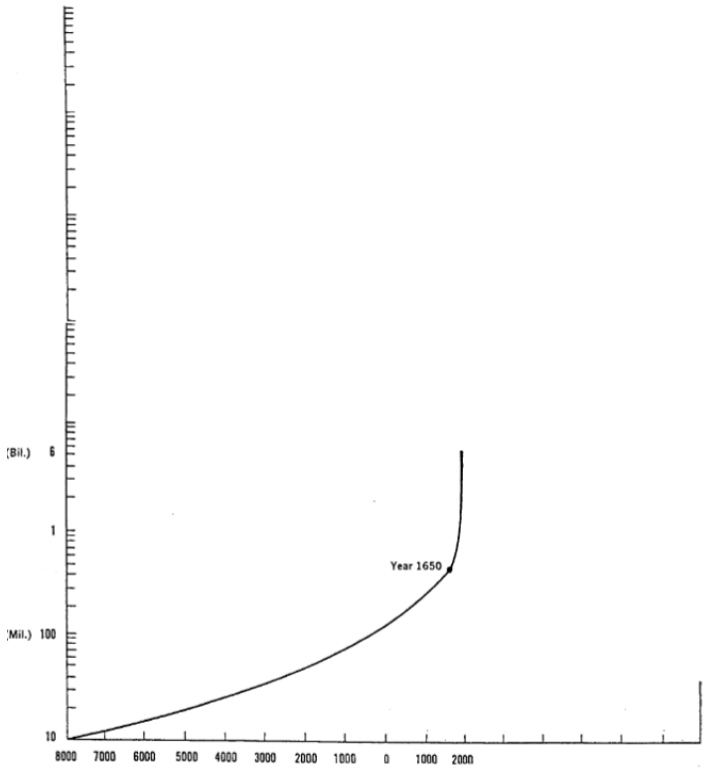

Figure 9

When the population line reaches the top of the graph, then there will be one square meter of the Earth's surface for each person. The marriage of science and technology was visible on a time scale of 10,000 years. There will be another event visible on a scale of 10,000 years. No one knows what that event will be; it depends on a number of things. Denial is the father of extinction, risk assessment is the mother of survival.

2. GODS

2.1 Agency – The Illiad

The Trojan war is said to have ended in 1184 BC. Homer's epic is dated between 850 and 630 BC. It is perhaps the greatest work in Western literature, often framed as the death of the classic hero (Achilles) and the birth of the modern hero (Odysseus) who prevails by trickery rather than valor. So great is the Illiad that everyone can find in it what they want to see. Here's what I see: Homer is an atheist and the Illiad is a searing indictment of religion and warfare glimpsing the transfer of agency from gods to men through a frail shred of good in humanity.

The book is mostly about the gods. They are vain, supercilious, cowardly, deceitful and thoroughly despicable. They are not worshipped but rather appeased, out of fear of retribution. History is just a game board for their amusement. Men attribute everything to the gods; if they act cowardly it's because the gods filled them with fear. If they take heart and prevail, the gods gave them that heart. Hercules killed his children because jealous Hera, to get even with Zeus for having screwed Hercules' mother Alcmene, made them appear to Hercules as monsters. Humans lack any agency. Homer mocks the gods from beginning to end. Here's one example which I haven't seen others pick up. Hera wants to seduce Zeus so she can help the Greeks against Zeus' favored Trojans (Book 14). She tricks Aphrodite into loaning her magic bra, and tricks Sleep by promising him her daughter. Zeus is aroused and courts Hera by telling her that he is hornier for her than for all the other women and goddesses he has raped. *"But for us twain, come, let us take our joy couched together in love; for never yet did desire for goddess or mortal woman so shed itself about me and overmaster the heart within my breast—nay, not when I was seized with love of the wife of Ixion, who bare Peirithous, the peer of the gods in counsel; nor of Danaë of the fair ankles, daughter of Acmsius, who bare Perseus, pre-eminent above all warriors; nor of the daughter of far-famed Phoenix, that bore me Minos and godlike Rhadamanthys; nor of Semele, nor of Alcmene in Thebes, and she brought forth Heracles, her son stout of heart, and Semele bare Dionysus, the joy of mortals; nor of Demeter, the fair-tressed queen; nor of glorious Leto; nay, nor yet of thine own self, as now I love thee, and sweet desire layeth hold of me."*

Zeus doesn't know he's being played.

The Illiad comprises just a few weeks of the 10 year war, starting with Agamemnon seizing Achilles' prize salve girl Briseis (won by killing her parents), with whom Achilles has fallen in love and promises to marry – not unlike the crime for which the Greeks attacked Troy (the gods made Paris and Helen do it). Achilles pouts, the Trojans prevail under Hector until Hector kills Patroclus. Achilles then kills Hector and drags off his body to feed to the dogs. The body of Hector is a prize greater than the fairest slave girl. Now Zeus asks Achilles to give up this prize and let Priam take the body for proper burial – for a proper ransom. Achilles accedes.

But when Priam visits Achilles, Achilles finally does something which the gods have not told him to do.

"*Then spake to him in answer swift-footed, goodly Achilles:"Thus shall this also be aged Priam, even as thou wouldest have it; for I will hold back the battle for such time as thou dost bid." When he had thus spoken he clasped the old man's right hand by the wrist, lest his heart should any wise wax fearful."*

Compassion! No god told Achilles to take Priam's hand. Finally someone acts on his own agency, free of the gods' manipulation. The Iliad ends with the burning of Hector's body and burial of this bones. It's frail hope but hope nonetheless. Hope for human agency, to do what?

2.2 Immediacy: Passez outre

Confronted with extra ordinary phenomena, the mind tries to place them by projecting beliefs and expectations. Jeanne d'Arc was extra ordinary in the extreme. George Bernard Shaw said that Jeanne's voices and visions prove nothing, "*but the variety of the conclusions reached show how little our matter–of–fact historians know about other people's minds, or even about their own*". Firm in the knowledge of his own mind, George concluded that "*Joan got a far fairer trial from the Church and the Inquisition than any prisoner of her type and in her situation gets nowadays in any official secular court; and the decision was strictly according to law*" (still smarting George?). Mark Twain considered his historical novel on Jeanne his best work. It is ignored today. Philosophers should study Jeanne's trial, for it raises questions more profound than those raised by the trials of Socrates, Giordano Bruno or Galileo.

History
The divine rights of Medieval kings were so tangled that god could make known his preference only through war. After the battle of Agincourt (1415) the English control most of France north of the Loire river, affirming Henry V's claim to the French throne through his maternal grandmother (maternal succession was not generally recognized). King Charles VI of France is insane and before dying in 1422, signs the Treaty of Troyes disinheriting his son Charles VII and giving the kingdom to Henry V, who dies before his coronation. On March 4, 1429 a peasant maid of (she thinks) 17 named Jeanne (no last name) tells Charles VII that she has been ordered by god to conduct him to his coronation in Reims. She is examined by the Archbishop of Reims (ecclesiastic superior of Pierre Cauchon see below) at Poitiers who finds her without blemish but recommends that she first give a sign. With troops from Charles VII she lifts the 6 month siege of the key Loire city of Orleans in 4 days. Before going into battle she orders all her soldiers on their knees to confess their sins; a sinful heart fears death. Jeanne never unsheathes her sword in battle but leads with her banner. A succession of stunning English defeats leads to Reims in 70 days where Charles VII is crowned. Charles VII wants to negotiate with the English but Jeanne adamantly insists on pursuing the war. While Charles VII dithers, Jeanne writes to the Hussites in Bavaria who question papal authority: "I shall destroy your empty and abominable superstition and strip you of either your heresy or your lives". For context, Jeanne wasn't the only visionary. Jeanne-Marie de Maillé , a woman of noble birth had visions and made prophecies – one concerning the multiple popes. In September 1430 a woman named Pieronne announced in Paris that she talked with god and that what Jeanne "did was well done and was God's will'. For refusing to recant Pieronne was burned at the stake. Nothing new, French mystic nun Marguerite Porete was burned in 1310 for refusing to abjure her previously approved book The Mirror of Simple Souls. The Czech priest Jan Hus, inspiration for Martin Luther, was burned at the stake for heresy on July 6 1415. His followers, the Hussites, defeated five consecutive papal crusades between 1420 and 1431. ans was burned at the stake on July 76, 1415.

At Compiegne the French are pushed back to their ramparts. Jeanne covers the retreat but the city gates suddenly close leaving her to be captured by the faux French Burgundians fighting for England. God's fickle preferences changed. The Burgundians expect a handsome ransom but Charles VII offers none and remains deafeningly silent. The archbishop of Reims reports that a new emissary from god, William the Shepherd, is with Charles' forces. The shepherd disappears in English captivity a few months later. The English don't want Jeanne to live but don't want her blood on their hands. They bankroll Burgundian Bishop Pierre Cauchon and pay a ransom equivalent to 1000 first class war steeds (the English paid Cauchon 100 equivalent steeds per year). Although Jeanne has already been investigated by his ecclesiastic superior, on 9 Jan 1431 Cauchon initiates an unauthorized ecclesiastical trial at Rouen whose verdict is never in doubt. After 5 months of interrogation Jeanne is convicted of dressing like a man and "listening to voices". Faced with incineration she confesses. Her voices then admonish her that she has sold her soul to save her body. She retracts her confession on May 30, 1431 and is burned at the stake. She is 19 years old.

On July 7, 1456, after taking possession of the transcripts at Rouen, the church reverses the 1431 verdict. Cauchon is excommunicated post mortem, his body is retrieved from his recent grave and tossed into a public sewer. On May 16, 1920, 500 years after her death, Pope Benedict XV canonized her making her the patron saint of prisoners, France, and soldiers – particularly those who are women.

Figure 1 Jeanne with sword (left) and Jeanne with banner (right)

Philosophy

Transcripts of her examination at Poitiers, her trial at Rouen and extensive eye witness testimony during her rehabilitation of 1456 are all available (the Book of the Poitiers examination is lost, but recounts of participants survive) . Here are a few snippets from her interrogation at Poitiers and Rouen:

- Asked why god needed soldiers "*In the name of God, the soldiers will fight and God will give victory."*
- Asked by a prelate from Limousin whether her voices spoke French: "*Better than you.*"
- Asked if she believed in God: *"Verily, more than you!"*
- Asked why they should believe her without a sign to support her claims: "*In the name of God, I have not come to Poitiers to give signs; but take me to Orléans, and I will show you signs of the purpose for which I am sent."*

At her trial at Rouen, Cauchon asked who advised her to dress like a man: "*Passez outre"*[15] – her answer to many of Cauchon's questions..

In all of human history there is no one remotely comparable to Jeanne, not Boudica, not Semirames, not Hyppolyta. A hundred years war with untold death and misery could not reveal god's regal preference. An illiterate penniless maid of 17 accomplished in 70 days what these armies could not. The church found her a heretic then a saint. According to their rules she is both (think about it). Her perfection of purpose was born of her apodictic experience of visions and voices. Apodictic experiences are self-validating, immediate and incontrovertible. The problem is that one cannot transfer them to others. Jeanne can tell others about her visions but she can't cause others to have them. Yet these are social constructs – had she heard voices of Hypatia and Sumayyah bint Khayyat instead of St. Catherine and St. Margret no one would have listened. She cannot transfer, she can only attest. If under threat of painful death she disavows her voices, then they are invented by Satan who is finally purged −or did she really disavow god to save her body? If she prefers death to abjuration the voices are from god−or did Satan prevail after all. At the end of her trial in 1431, asked if her voices were good or evil spirits she replied *"Soint bons, soint mauvais esperits, ilz me sont apparus."* God or

[15] No simple English equivalent, suggestions include: *bypass, carry on regardless, circumvent, dispense with, disregard, ignore, go beyond, get over it, get past, move on, override, overcome, overlook, overrule, skip,....*

Satan, its irrelevant. Jeanne is not attesting to the truth of her visions…she doesn't care. Yet she still chose death on the pyre. To what was she attesting? This question is Jeanne's contribution to philosophy.

Plato, channeling Socrates, based the highest knowledge on direct intuition of the Good. Like Jeanne's visions such intuition is apodictic, self-validating and non−transferable. Plato never asked if his Good was really good. Why should he? That's the point of apodicticity. Otherwise he would need a meta−intuition certifying his intuition...ad infinitum. Plato, in the person of Socrates, attested. Poet, philosopher and early heliocentric advocate Giordano Bruno was imprisoned in Rome during his 7 year trial for heresy. He recanted but then repudiated his recantation and was burned at the Campo de 'Fiori in 1600. He attested. Galileo was also convicted of the heliocentric heresy in 1633. Rather than attest he publicly denied that the Earth moved. His truth didn't require his martyrdom, just his telescope, which is transferable. Galileo's conviction was annulled in 1992 after 359 short years. In 2000 Pope Paul II apologized for burning Bruno, without reversing their verdict of heresy.

The immediacy of Jeanne's visions is unlike the philosopher's putative intuition of truth. Did St Margret and St Catherine really speak to her? Did they exist? Both are probably inventions. The point lies elsewhere, move on, get over it. We still can't place Jeanne. Mark Twain's effort is the best so far IMO. Maybe novelists are better at transferring experiences than "matter−of−fact" historians.

What ended the Middle Ages? Some say the fall of Constantinople in 1453, some say the Gutenberg Bible in 1455, some say the rehabilitation of Jeanne d'Arc in 1456 when the world began to understand *Passez Outre*.

2.3 Lying by Omission: Islamic Pluralism[16]

How many Christians and Jews know that the father of Jesus repeatedly issued instructions like: "*Now therefore kill every male among the little ones, and kill every woman that hath known a man by lying with him. But all the women children that have not known a man by lying with him, keep alive for yourselves"* (Moses, **Numbers, 31.17/18**). *"thou shalt save alive nothing that breatheth"* (Moses, **Deuteronomy 20.16**). *"And thou shalt do to Ai and her king as thou didst unto Jericho and her King"* (God to Joshua **Joshua. 8 :2**). A commander issuing such orders would have been hanged at Nuremburg. It all springs from*"….for the Lord, whose name is Jealous, is a jealous god"(* **Exodus** 34:14). Most do not know that Jesus said things like "If a man abide not in me, he is cast forth as a branch...men gather them, and cast them into the fire, and they are burned "(Jesus, **John 15;6**). The Koran is much worse, "*We renounce you: enmity and hate shall reign between us until you believe in Allah only*"(**Koran 60:2**) *"when the sacred months are over slay the idolaters wherever you find them"* (**Koran 9:5**) *… "make war upon them until idolatry shall cease and God's religion shall reign supreme."* (**Koran 8:40***)…"Kill them wherever you find them"* (**Koran 2 :193***)… "Prophet, make war on the unbelievers… Fight against them until idolatry is no more and Allah's religion reigns supreme….Idolatry is worse than Carnage" (**Koran 2:191-193**)."* (**Koran 9:73**). *"The only true faith in God's sight is Islam."* (**Koran 3:19**). Most believers don't know this because of dissimulation, lying by omission. Lying by omission IS lying. [17]

The Centre for Islamic Pluralism's (CIP) recent *A Guide To Shariah Law And Islamist Ideology In Western Europe* (2007-2013) by Stephen Suleyman Schwartz, Dr Irfan Al-Alawi is a good book. It is not a stalking horse for islamists and does not pander to the Muslim victim narrative. A guide to Shariah law it is not, but it is a good lense for focusing on the dissimulations of modern religious worship. Indeed, dissimulation is a hallmark of religious *worship*. The ancient Greeks (see Chap 2.1) did not dissimulate about their gods because they didn't worship their gods, they *appeased* them, knowing how terribly they could behave.

The message of *A Guide To Shariah Law* is that Islam contains traditions which reject Shariah jurisprudence in favor of their "traditional" laws and mores. European governments should encourage this "pluralism" as it offers more space to "secular Muslims" (??) and Muslims of Western orientation. Al Sistani's A Code of Practice For Muslims in the West is offered as a "striking expression of the principle of acceptance of Western law by traditional Muslim immigrants" (loc 919). If you're wondering why asking Muslims to obey the law is "striking" read on. In any event, the vice squad might take issue with temporary marriages: "The formula for solemnizing the temporary marriage is as follows: The woman says to the man: 'I give myself to you in marriage for the dowry of (x) for the time period (x).' The man immediately says, 'I accept the marriage.' " Differences in the Muslim populations of Britain (Pakistanis), Germany (Turks, Kurds, Alevis) and Spain (Berber, Sufi, Amazigh) are related to differences in assimilation. Many of these groups reject the Wahabi call for global jihad. There are many nuggets. We learn eg that Wahabi hospital personnel in Britain refuse to wash their hands with alcohol, using soap instead which doesn't disinfect.

The naïve reader comes away with the impression that this personal Islam has been hijacked by the Wahabi "cult" from Saudi Arabia. The reader can judge whether CIP is lying by omission only if (s)he knows what is left out.

---

[16] Based on Amazon book review Tepid, Timid and Tragic, https://www.amazon.com/SHARIAH-ISLAMIST-IDEOLOGY-WESTERN-EUROPE-ebook/dp/B00B5GW30S/ref=sr_1_1?crid=2FFN4LJ65S7PM&dib=eyJ2IjoiMSJ9.i557pARCbG2fABfklCLQ3Q.sEx6KouNZGMiICKoBvU1pHtYxbMVsYJ-sEuS5hFncBc&dib_tag=se&keywords=A+GUIDE+TO+SHARIAH+LAW+and+ISLAMIST+IDEOLOGY+in+WESTERN+EUROPE&qid=1723891176&sprefix=a+guide+to+shariah+law+and+islamist+ideology+in+western+europe%2Caps%2C265&sr=8-1#customerReviews

[17] The theme is further explored in a companion Amazon review " 3=7?" https://www.amazon.com/product-reviews/1250135133/ref=acr_dp_hist_1?ie=UTF8&filterByStar=one_star&reviewerType=all_reviews#reviews-filter-bar

As traditions substituting their own customary law for Shariah, CIP mentions Suleyman the Magnificent, and the 13th century Mongol invaders. Suleyman's traditions featured massive "child collection" (*devşirme*,) or enslavement of young Christian boys from the Balkans and Anatolia, conversion to Islam and conscription into military service. The "Sword of Islam", Sufi Mongol Tamerlane, led a jihad that is credited with murdering 17 million people (5% of the world population) (https://en.wikipedia.org/wiki/Timur). CIP cites 5 sources of "traditional Islam", including
Reliance of the Traveller: A Classic Manual of Islamic Sacred Law festooned with references to the Qur'an and Hadiths, comments by different schools of Islamic jurisprudence, and 1991 certified English translation. It is said to represent the consensus of 75% of the Islamic religious scholars. Bear with me, the message is the volume (the letters N, A, O etc refer to commentators).

Jihad *"It is offensive to conduct a military expedition against hostile non-Muslims without the caliph's permission (though if there no caliph no permission is required)…The caliph makes war upon Jews, Christian and Zoroastrians (N: provided he has first invited them to enter Islam in faith and practice, and if they will not, then invited them to enter the social order of Islam by paying the non-Muslim poll tax"…. "The caliph fights all other peoples until they become Muslim (O; because they are not a people with a Book, nor honored as such, and are not permitted to settle with paying the poll tax (jizya)) (N: though according to the Hanafi school, peoples of all other religions, even idol worshippers, are permitted to live under the protection of the Islamic state, if they either become Muslims or agree to pay the poll tax, the sole exceptions to which are apostates from Islam and idol worshippers, who are Arabs, neither of whom has any choice but becoming Muslim" (p. 603)*

Jesus *"After his final coming, nothing but Islam will be accepted from them, for taking the poll tax is only effective until Jesus' decent .... The coming of Jesus does not entail a separate divinely revealed law, for he will rule by the law of Muhammad " (p. 602)..*

Apostasy *"Leaving Islam is the ugliest form of unbelief (kufr) and the worst….When a person who has reached puberty and is sane voluntarily apostatizes from Islam, he deserves to be killed"*
"Acts that entail leaving Islam: ….
(7) to deny any verse of the Koran or anything which by scholarly consensus belongs to it, or to add a verse that does not belong to it"
(17) "to believe that things in themselves or by their own nature have any causal influence independent of the will of Allah
(18) to deny the existence of angels or jinn, or the heavens
(19) to be sarcastic about any ruling of the Sacred Law
(20) to deny that Allah intended the Prophet's message...to be the religion followed by the entire world.. (p.595)

Forced marriages *"Whenever the bride is a virgin, the father or father's father may marry her to someone without her permission, though it is recommended to ask her permission if she has reached puberty. A virgin's silence is considered as permission" (p 522), … " A guardian may not marry his prepubescent daughter to someone for less than the amount typically received as marriage payment by similar brides, nor marry his prepubescent son to a female who is given more than the amount typically received. (p.533).*

Genital mutilation *"Circumcision is obligatory (O: for both men and women. For men it consists of removing the prepuce from the penis, and for women, removing the prepuce(Ar.bazr) of the clitoris (n: but not the clitoris itself, as some mistakenly assert).(A: Hanbalis hold that circumcision of women is not obligatory but sunna, while Hanafis consider it is a mere courtesy to the husband)." (p.59)*

Slavery *[the passages on slavery were not translated into English, the following are taken from "Rules of Warfare"]"When a child or a woman is taken captive, they become slaves by the fact of capture, and the woman's previous marriage is immediately annulled." ..."When an adult male is taken captive, the caliph considers the interests (O; of Islam and the Muslims) and decides between the prisoner's death, slavery, release without paying anything, or ransoming himself in exchange for money, or for Muslim captives held by the enemy. If the prisoner becomes a Muslim (O before the caliph chooses any of the four alternatives) then he may not be killed, and one of the other three alternatives is chosen"(p.604)*

Unlawful knowledge, music, pictures *"Unlawful knowledge includes (2) Philosophy...(5) the sciences of the materialists. (p.14) ... 'Allah mighty and majestic sent me as a guidance and mercy to believers and commanded me to do away with musical instruments, flutes, strings, crucifixes and the affair of the pre-Islamic period of ignorance' ... All of this is explicit and compelling textual evidence that musical instruments of all types are unlawful" (p.775). [p44.1 (1) "Every maker of pictures will go to the fire, where a being will be set upon him for each picture he made, to torment him in hell"(p.683)*

The litany could go on and on. Is this "Traditional Islam" ? If not, where are the fatwas forbidding all these things. Which of the "plural Islams" is compatible with the 1948 UN Universal Declaration of Human Rights? ("No one shall be held in slavery… Marriage shall be entered into only with the free and full consent of the intending spouses… Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief…") . Which Islams have repudiated the 1990 Cairo Declaration of Human Rights in Islam ("it is prohibited to take away life except for a Shari'ah prescribed reason… Men and women have the right to marriage, and no restrictions stemming from race, color or nationality [not religion??] shall prevent them from enjoying this right… Everyone shall have the right to express his opinion freely in such manner as would not be contrary to the principles of the Shari'ah…. All the rights and freedoms stipulated in this Declaration are subject to the Islamic Shari'ah").What purpose could CIP serve by giving a grossly distorted picture of Islam? "Tepid, Timid" is a charitable interpretation…the authors just want to perform their rituals in peace. We move on to the Tragic.

The word "Qur'an" occurs in this book 28 times, but the Qur'an is NEVER cited or referenced by verse. Compare this to Jihadi texts like the glossy magazine Inspire, which is laced with citations. Jihadis apparently think the non-abrogated Qur'an verses work for them. Killing apostates and genital mutilation are from the Hadiths, not the Qur'an, but slavery (8:69, 24:33,23:1, 4:1, 4:25, 33:51) , hatred of unbelievers (60:2, 3:24, 4:141, 3:117, 60:1, 9:19, 5:49, 5:57, 3:85, 3:19, 4.88-89) wife beating (4:34) and genocide (a short list of verses: 2:190-3, 2:217, 2:214, 8:40, 9:4, 9:73, 9;122, 47:3, 66:9) certainly are (I use the definition of genocide in the 1948 UN convention on the prevention and prosecution of the crime of genocide). Lets not omit Allah's revelation that the sun sets in the west in a pool of black mud (18:86). Islamic apologists are fond of saying that one can't just grab sentences from the Qur'an without interpretive context. Guess what, the Qur'an has already been interpreted, for 1400 years. Did the all jurists quoted in the Reliance get it all wrong? for 1400 years? The prophet married Aisha when she was 6 and consummated the marriage when she was 9. How can a Muslim repudiate forced marriage without repudiating the verse "You [Muhammad] may have whomever you desire; there is no blame"? (33:51). The "scholarly consensus" regarding slave girls was rendered by Muslim scholars at Al Azhar who in 1994 found linguist Nasr Abu Zayd guilty of apostasy inter alia for his "denunciation… of the permissibility of the ownership of slave girls, a principle considered religiously proven without doubt" (p. 16 of the judicial opinion, https://en.wikipedia.org/wiki/Nasr_Abu_Zayd .) CIP cannot repudiate the Reliance without committing apostasy. Instead they dissimulate, saying only that Muslims shouldn't practice slavery, forced marriage, genocide etc. when living in a non-Muslim country.

What legitimizes authority? Divine Revelation? the Will of the Strongest? Natural Right? Social Contract, Custom/Tradition? Muslims can't appeal to Divine Revelation without somewhere citing the Qur'an. Worshipping a god who condones slavery and rape takes Natural Right and Social Contract off the table. Are the predations of a Suleyman or a Tamerlane legitimate because they were the strongest? No? That leaves only tradition and herein lies the tragedy. Muslim immigrant children are losing their traditions with nothing to replace them. They are asking why are Islamic countries backward, brutal and weak? Why do so many flee to the West, while Westeners aren't flocking to the Middle East? Why are Muslims slaughtering each other? Looking for answers in the Koran, the Hadiths, the Reliance etc, they read passages like those referenced above and turn to their elders saying: 'you call yourself a Muslim, but you're not doing what god commands:' 'Islam is weak and disrespected because you follow the laws of men and not the laws of god'. The Centre for Islamic Pluralism has no answer, now Islam is collecting their children. The alternative starts by seeing the UN Universal Declaration of Human Rights not as a concession but as a disideratum.

2.4 Midrash: When Trading Fails

The spirit of Capitalism is that 'everything, but then really *everything,* should be calculated in terms of capital' (chapt 1). Capital is liquid, it is traded. Imagine this discussion between the Pope and the Caliph:

"Ok, We'll give up the virgin birth if you give up Jihad"
"No deal, you'll have to throw in the resurrection"
"Well, in that case you'll have to give up Jesus returning as a Muslim"
"Deal".

Can't happen. Religious conflicts never resolve.

My student Jan Norstrom pinpointed the defining feature of dogmatic religion: you can't trade it. Received wisdom distinguishes revelations and myths. Revelations are apodictic, immediate, incontrovertible and self-validating. Myths don't pretend to be true, they are fancies intended to represent values. Is Hercules a myth? Santa Claus? Moses? Jesus? Mohammad? George Washington?

Between the twin continents of Myth and Revelation lies a third continent shrouded in mist, its name: MIDRASH. Mist does not comport with clear definitions. According to the Encyclopedia Britannica, *"Midrash was initially a philological method of interpreting the literal meaning of biblical texts. In time it developed into a sophisticated interpretive system that reconciled apparent biblical contradictions, established the scriptural basis of new laws, and enriched biblical content with new meaning."*

Modern Hebrew scholars describe it thus:
"They reimagine dominant narrative readings while crafting new ones to stand alongside—not replace—former readings. Midrash also asks questions of the text; sometimes it provides answers, sometimes it leaves the reader to answer the questions"[18]

"[It is] a Jewish mode of interpretation that not only engages the words of the text, behind the text, and beyond the text, but also focuses on each letter, and the words left unsaid by each line"[19].

I propose a sharper philosophical definition: *Midrash is helping gods express themselves*. It captures gods' meaning and thus enjoys revelatory status. Were gods to disagree they would surely intervene. A prosaic rendering is: throwing stuff out and adding other stuff to clarify god's meaning. Dissimulation aims to conceal the intended meaning, midrash aims to reveal it. Anyone can do it. Here is an account from one of my favorite Fallen Angels:

"Although the Exodus myth had been told considerably earlier—as allusions in Hosea, Amos, and Micah, dating from the eighth century BCE, make abundantly clear—the era of its literary elaboration and cultic institutionalization only dawned in the sixth century BCE, the period of Babylonian captivity. In particular, its great moment came with the return from exile, when "Israel" had to be reinvented as an ethnic and religious identity and established on the basis of a political, social, and religious constitution. With the help of the Exodus story, those faced with this task succeeded in creating a memory that defined them as a group, anchoring them in the depths of time while also committing them to a common future. What they were doing

[18] *Gafney, Wilda (2017). Womanist Midrash : a reintroduction to the women of the Torah and the throne (First ed.). Louisville, Kentucky. ISBN 9780664239039. OCLC 988864539.*
[19] *Lovelace, Vanessa (2018-09-11). "Womanist Midrash: A Reintroduction to the Women of the Torah and the Throne, written by Wilda C. Gafney". Horizons in Biblical Theology. **40** (2): 212–215.*

was more than just history-writing; they were declaring their allegiance to an identity, fashioning a collectively binding self-definition in the medium of narrative and memory. In the two forms of storytelling and lawgiving, the narrative and the normative, the Book of Exodus codifies the one all-transforming, truly epochal revelation in which God emerged from his inscrutable concealment—for the Jews, once and for all; for Christians and Muslims, for the first time—to manifest his will to his people, so establishing a completely new relationship to the world, to time, and to the divine". Assmann, Jan. The Invention of Religion: Faith and Covenant in the Book of Exodus (p. 5). Princeton University Press. Kindle Edition.

Story? Myth? Creating a memory? Creating a memory! isn't that called lying, and brainwashing? Isn't Assmann saying that the authors of Exodus were just making it all up? The crusades, jihads, inquisitions, pyres…all based on "created memories". Was god's "epochal revelation" that "*all the women children that have not known a man by lying with him, keep alive for yourselves"* (Numbers, 31.17/18) so different from Zeus exposing himself to Hercules' mother Alcmene, to whom Hera took such umbrage? But wait, doesn't everyone already have an identity, as human being? What's wrong with that?

Islam takes midrash to the next level: "*When We change one verse for another (Allah knows best what He reveals), they say: 'You are an impostor'.... Say: 'The Holy Spirit brought it down from your Lord in truth to reassure the faithful, and to give guidance and good news to those that surrender themselves to Allah'.(Koran 16:103). "Whatever communications We abrogate or cause to be forgotten, We bring one better than it or like it. Do you not know that Allah has power over all things?" (Koran 2: 106)*

Native American (Huron-Wendat) philosopher Kandiaronk (1625–1701) reflects on god's ability to express him/herself: "*For myself, I've always held that, if it were possible that God had lowered his standards sufficiently to come down to earth, he would have done it in full view of everyone, … . Then we would all have had exactly the same religion, …. Instead, there are five or six hundred religions, each distinct from the other, of which according to you, the religion of the French, alone, is any good, sainted, or true*." Graeber, David. The Dawn of Everything: A New History of Humanity (p. 53). Farrar, Straus and Giroux. Kindle Edition. Kandiaronk's god does not consecrate tribal identities. Just imagine what will be possible once humankind lifts its curse on itself?

Historicity is irrelevant for myths, who cares whether Hercules or the Cyclops really existed. Revelations are different. Historicity cannot augment revelations' apodictic certainty, but they're useful in proselytizing. One doesn't obtain apodictic certainty by incrementally removing uncertainty. One is – pick your word – brainwashed, groomed, re-programmed, purified - to leap to apodictic certainty. Proselytizing does the grooming. If the historisticy bar is set low enough to admit the story in Exodus, then many Egyptian texts also clamor for admission. However, *"If, given their clear political and didactic tendency, we have reason to doubt that the Egyptian lamentations have any basis in historical fact, then such skepticism is all the more warranted when it comes to the motives informing the account of the Egyptian plagues found in Exodus".* *Assmann, Jan. The Invention of Religion: Faith and Covenant in the Book of Exodus (p. 154). Princeton University Press. Kindle Edition.* Jesus and Paul wouldn't fare much better[20]. Even Muhammad, 'born in the clear light of history' has serious challengers[21]. George Washington's historicity is not in play, but some might midrash reports that after the revolution he stood up and left church without taking communion[22] to show he was not a Christian. To my knowledge, no one has tried to

[20] See for example "First, the closer scrutiny the Pauline texts receive, the clearer it becomes (and by now it seems mighty clear indeed) that the epistles present us with many of the same challenges the Gospels did. … In short, the historical Jesus problem replicates itself in the case of Paul. …" Price, Robert M. The Amazing Colossal Apostle: The Search for the Historical Paul . Signature Books. Kindle Edition. "What our theological teachers taught us about the impossibility of knowing the historical Jesus (concerning whom we know nothing more than the fact that there had been such a person) as well as about the creative imagination of the Christian communities after Jesus was shocking for many students." Detering, Hermann. The Fabricated Paul. Early Christianity In The Twilight. . Kindle Edition.

[21] Holland, Tom (2012) In the Shadow of the Sword: The Birth of Islam and the Rise of the Global Arab Empire

[22] https://en.wikipedia.org/wiki/Religious_views_of_George_Washington

midrash Jeanne and her historicity is as good as it gets in the 15th century. Moses, Jesus and Muhammad would be very jealous indeed of the transcrpts of her trial at Rouen.

3. Certainty

Certainty is knowledge, unanimity, power, authority. From Plato's direct intuition of the Good, to Aristotle's syllogisms, to Aquinas' revelation, to Descartes' clear and distinct ideas, to Kant's synthetic a priori statements, to Wittgenstein's atomic propositions, to the axioms of logic and set theory, philosophers have vainly pursued certainty. Yet it eludes us. Lets start from from something which doesn't elude us: UNcertainty, that which disappears as we become certain. Notice this: lying always claims certainty. In the language of uncertainty, lying is impossible: '*I believe with subjective probability $x < 1$ that I didn't have sex with that woman'*. Obfuscation on the other hand is possible, and flourishing. Nowhere is this more evident than at the science / policy gradient.

Most people reason under uncertainty by first constructing a deterministic narrative assuming all uncertain quantities take their "nominal" or "most likely" or "expected" values and then flagging uncertainties by sprinkling the narrative with uncertainty qualifiers like "most likely that", "reasonable to expect that', 'generally agreed that', etc. This may be adequate for deciding whether to take an umbrella to work, but surely not for crafting policy to meet existential but uncertain challenges like Climate Change. The alternative to hurling words at uncertainty is to quantify it through a process of rational consensus based on empirical measures of performance. Subsequent chapters elaborate this approach. The next chapter is devoted to uncertainty and its quantification.

3.1 Partial Belief and Messaging Climate Change[23]

The US National Research Council's 2010 report, *Advancing the Science of Climate Change* illustrates reasoning and communicating under uncertainty. Using the calibrated uncertainty language of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4, 2005), the NRC report bases its first summary conclusion on "high confidence" (at least 8 out of 10) or "very high confidence" (at least 9 out of 10) in six (paraphrased) statements:[24]

1. Earth is warming.
2. Most of the warming over the last several decades can be attributed to human activities.
3. Natural climate variability ... cannot explain or offset the long-term warming trend.
4. Global warming is closely associated with a broad spectrum of other changes.
5. Human-induced climate change and its impacts will continue for many decades.
6. The ultimate magnitude of climate change and the severity of its impacts depend strongly on the actions that human societies take to respond to these risks.

Logicians refer to the domain of everyday discourse as the "natural language", where rules of reasoning are not rigorously defined. The IPCC hoped to raise the debate on climate uncertainty by injecting precisely defined uncertainty qualifiers into the natural language. The evident problem with this approach is that the propagation of uncertainty through a chain of inference is conducted in the natural language. Indeed, what is the confidence that all six statements hold? It is not even clear whether "*all statements have a 0.8 chance of being true*" means "*each statement has a 0.8 chance of being true*" or "*there is a 0.8 chance that all statements are true*". The natural language obscures the gaping difference between these two statements. Attempting a rigorous reconstruction of the above chain of inference highlights the limitations of uncertainty propagation in the natural language.

Consider the second statement. Does it impute high confidence to 'Earth is warming AND humans are responsible' ("THE warming"), or to the conditional statement 'GIVEN that the Earth is warming, humans are responsible'? These are very different statements, and again, the natural language masks this difference. Since the Earth's warming is asserted in the first statement, perhaps the latter, conditional, statement is meant (if, after months of word-smithing, the authors knew what they meant, we wouldn't be asking). In that case, the likelihood of both statements holding is the product of their individual likelihoods: chance of the condition (Earth warming) × chance the conclusion based on the condition (Humans caused it). If the first two statements enjoy "high confidence", then both can hold with only "medium confidence" ($0.8 \times 0.8 = 0.64$). If the remaining 4 statements are treated similarly, a high confidence in all of them becomes a paltry ($0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.8 =$) 0.26.

The calibrated language translates "virtually certain" as 99%–100% probability (Mach et al 2011, Mastrandrea et al 2010). Suppose the US Nuclear Regulatory Commission licensed nuclear reactors based on the finding that each reactor's safety each year was "virtually certain". With 100 commercial nuclear reactors, each with a probability of 1/100 per year of a meltdown,.... well, do the math. That is the point: to propagate uncertainty you may have to do some math. You can't do it by the seat of your pants, you need to think probabilistically

[23] This chapter abridges Cooke, Roger M. (2015) "Messaging climate change uncertainty with Supplementary Online Material" Nature Climate Change 5, 8–10 (2015) doi:10.1038/nclimate2466 Published online 18 December 2014, http://www.nature.com/nclimate/journal/v5/n1/full/nclimate2466.html

[24] See National Research Council (2010, 4–5) http://www.nap.edu/catalog.php?record_id=12782): "As discussed in Appendix D, high confidence indicates an estimated 8 out of 10 or better chance of a statement being correct, while very high confidence (or a statement than an outcome is 'very likely') indicates a 9 out of 10 or better chance."

and that involves numbers. The calibrated language has the important virtue of making problems of uncertainty propagation in the natural language obvious, though apparently not obvious enough.

In more recent publications, IPCC eschews the numerical interpretation of confidence levels, calling them instead *"qualitative measure[s] of the validity of a finding, based on the type, amount, quality and consistency of evidence (e.g., data, mechanistic understanding, theory, models, expert judgment) and the degree of agreement" (AR6 p169) ie* BOGSAT (see chapter 4.3). Arguably the most important climate parameter is Equilibrium Climate Sensitivity (ECS), the amount by which Earth's surface temperature will eventually rise as a result of doubling the atmospheric concentration of CO2. In communicating the uncertainty surrounding ECS, likelihood is still interpreted probabilistically, but likelihood statements themselves are rendered with confidence levels. This leads to orgies of obfuscation like:

"ECS is extremely unlikely [< 5%] less than 1°C (high confidence), very unlikely [< 10%] greater than 6°C (medium confidence)." (AR5 p.1111)

Climate authority David Archer writes "*A climate change of the magnitude of the deglaciation, 5-6°C, would be catastrophic to human civilization* [25]". The above IPCC statement says *This catastrophe has a probability less than 10% of occurring and we have medium confidence in that*. Why only median confidence at the high end? How much would that 6°C have to rise to garner the same high confidence as the low end? AR5 and AR6 run over 5,949 pages. AR6 was approved by governments of 195 countries and no one asked this simple question. Why? Who is to blame? Answer: everyone[26].

The greatest barrier to communicating uncertainty is not some deficiency of the target audience; it is a deficient understanding of uncertainty on all sides. The logic of partial belief is subjective or Bayesian probability. Specialists have known how to "do" uncertainty for a long time, and it involves specialist training. You can't do it just by throwing words at it, as the National Research Council has amply shown. Not surprisingly, uncertainty has become a key part of the climate messaging: deniers and contrarians use uncertainty to shift the proof burdens.

"…we don't know what's causing climate change on this planet…. spending trillions and trillions of dollars to try to reduce CO2 emissions is not the right course for us." **Mitt Romney** 2011

Alarmists focus on worst cases to frighten us into impetuous action, and science messagers, crafty or clumsy, inflate their certainty. The IPCC obfuscates. The way forward starts with getting the uncertainty narrative right.

Back to the past

In 1977, under prodding from E.A. Feigenbaum (1977), the artificial intelligence community shifted attention from computer chess to "expert systems." Studying the strategies and heuristics of "grand masters" of science, they concluded that the grand masters did not reason probabilistically, and through the 1980s explored "alternative representations of uncertainty" (Cooke 1991), including certainty factors, degrees of possibility, fuzzy sets, belief functions, random sets, imprecise probabilities, and nonmonotonic logic, among many others. The proceedings of the premier conference "Uncertainty in Artificial Intelligence" have been digitized since

[25] Archer, David. The Long Thaw: How Humans Are Changing the Next 100,000 Years of Earth's Climate (Science Essentials) (p. 95). Princeton University Press. Kindle Edition

[26] Including me. As Lead Author for the chapter on Risk and Uncertainty in AR5, I diluted my critique for fear of being 'de-platformed', as I had been on previous occaisions (not with IPCC). Learning how and when to attest; a non-trivial problem.

1985 and provide a unique record of the development of alternative representations of uncertainty. Figure 1 shows the relative word fragment count of various approaches in 1985. The largest component is "belief function," followed by "Bayes," "fuzzy," and "certainty factor." Bayes is a proxy for subjective probability, it accounts for 26% of the total.

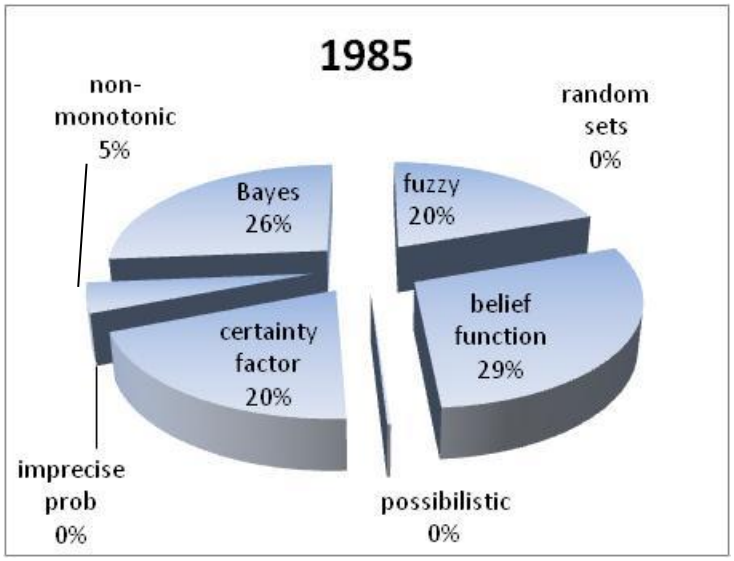

Figure 1: Word fragment counts in Uncertainty in Artificial Intelligence, 1985

In 2000 the balance has shifted; "Bayes" now accounts for 79% of the count. By 2012 the count is 97% "Bayes".

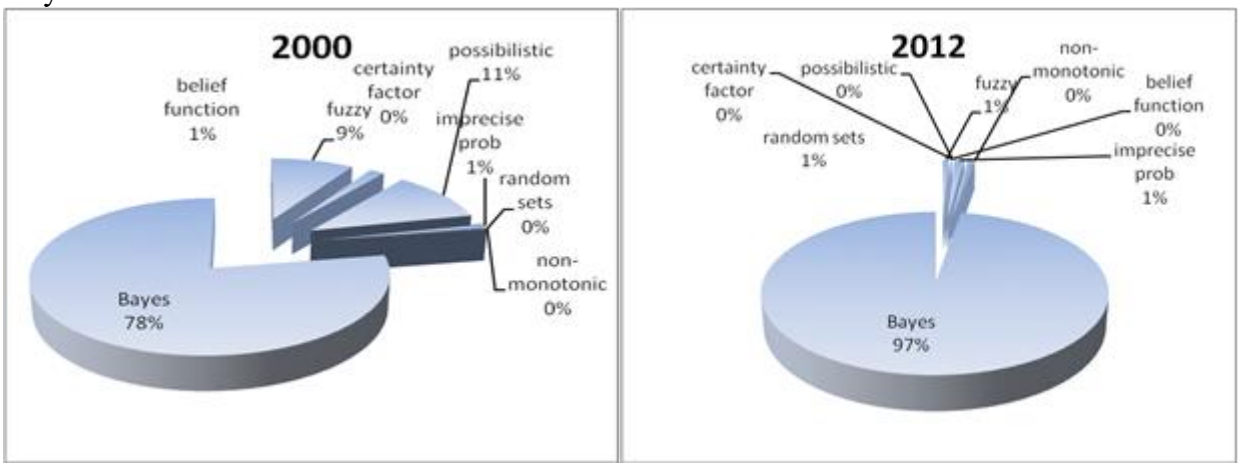

Figure 2. Word Fragment Counts in Proceedings of Uncertainty in Artificial Intelligence, 2000 (Left) and 2012 (Right)

Climate change is the current theater of alternative uncertainties. The following is a high pass over the representation of uncertainty in climate science, reducing technical jargon to the minimum. The logic of partial belief is rehearsed at the 30,000-foot level. Imprecise probabilities, deep, and Knightian uncertainties are overflown. For some uncertainties, the apparent depth may result from the lack of operational meaning; this is renamed shallow uncertainty in the next chapter.

The Logic of Partial Belief

The subjectivist interpretation of probability goes back to reverend Thomas Bayes in 1763, but the operational measurement of partial belief as subjective probability dates from Ramsey (1931). This brief section rehearses the best modern rendering; namely, that of Savage (1954).

Suppose that receiving $10,000 is better than receiving $100 in all situations. The event *France wins the next soccer World Cup* is "qualitatively more probable than" *USA wins the next soccer World Cup* for a subject if and only if (s)he prefers the lottery,

$10,000 if France…; $100 otherwise
to
$10,000 if USA…; $100 otherwise,

and if the relationship is additive:

$10,000 if France OR Belgium…; $100 otherwise
is preferred over
$10,000 if USA OR Belgium…; $100 otherwise.

Other axioms require that the qualitative probability relation is transitive, that multiplying rewards by a positive constant does not change preferences, and that there is a sufficient number of disjunct events. If your preference complies with these axioms then there is a unique probability measure P such that event A is qualitatively at least as probable as event B if and only if $P(A) \geq P(B)$. Ramsey's initial theory used preference equalities between "certainty equivalents" and gambles, and has led to the persistent misconception that subjective probability is restricted to betting. This is not true, and the step from a qualitative to quantitative probability is quite small. It can be shown that any event can be split into two sub-events of equal qualitative probability. It follows that we can make arbitrarily large uniform partitions. Let $A_1 \ldots A_n$ be a partition such that for every i,j, A_i is qualitatively as probable as A_j. Suppose that event X is qualitatively more probable than the union of the first k of the A's and qualitatively less probable than the union of the first $k+1$; then the probability of X is approximately k/n, and as n gets large, the numerical probability is uniquely determined.

Probability in this sense is the logic of partial belief. Logic does not tell us what to believe, it tells us, for example, that if we believe both A and *If A then B* then we should also believe B. Logic is normative; it says how we *should* transfer belief. It does not describe how people actually reason. Indeed, most people think that the following is a valid argument:

Only the fittest species will survive.
Cockroaches are the fittest species.
Therefore,
Cockroaches will survive.[27]

However, reasoning under uncertainty is VERY different than reasoning with certainty. Arguments which are valid if everything is certain are not "almost valid" once the tiniest bit of uncertainty leaks in. Consider this argument:

[27] Among third-year mathematics students at the Delft Technical University, 80% judged this a valid syllogism. To see that it is not valid, compare: "Only women get pregnant. Maria is a woman. Therefore, Maria gets pregnant."

Almost certainly all men are animals
Almost certainly all animals are mortal
Therefore *??*
Almost certainly all men are mortal

The premises and conclusions are true, but is the argument valid? Can the conclusion be false when the premises are both true? Consider:

Almost certainly all men weighing over 600 kg have less than 50 letters in their last name
Almost certainly all men having less than 50 letters in their last name weigh less than 600 kg
Therefore
Almost certainly all men weighing over 600 kg weigh less than 600 kg.

The premises are true and but conclusion is false, indeed, it is a contradiction.

As noted, most people, including scientists at the IPCC reason as if things were certain and then "account for uncertainty" by sprinkling the narrative with uncertainty qualifiers, "very likely, "high confidence", "virtually certain" etc. This is not correct, not even close. Probabilistic reasoning is much harder than reasoning with certainty and people commit errors in probabilistic reasoning in droves, even people at the top of the scientific food chain who set examples for others. For example, when told that "Bill is dull and good in math" most believe that "Bill is a jazz musician and an accountant" is more likely than "Bill is a jazz musician" (Tversky and Kahneman 1982).

The paradoxes of Allais (1953) and Ellsberg (1961) describe choice situations in which many people exhibit behavior violating the axioms guaranteeing the representation of partial belief as subjective probability. McCrimmon (1968) found that business executives willingly corrected violations of the axioms when made aware of them. Other authors (Kahneman and Tversky 1979; Schmeidler 1989; Quiggin 1993; Wakker 2010) account for paradoxical choice behavior by transforming the probabilities of outcomes into "decision weight probabilities," which play the role of likelihood in computing optimal choices but do not obey the laws of probability. Wakker (2010, p. 350) notes that decision weighting also fails to describe some observed behavior patterns.

To be sure, the representation of partial belief as subjective probability has limitations.

1. First and foremost, partial belief is *personal*, that is, it pertains to an individual. If a group of individuals' partial beliefs satisfy certain mathematical constraints, then by jointly observing various phenomena and updating their beliefs on the results of observations, members of the group will converge to a common probability distribution about those phenomena. Absent this "natural" convergence mechanism, differing partial beliefs are unavoidable.
2. Just as propositional logic does not capture every valid passage of thought, subjective probability does not capture all valid reasoning about partial belief. Events which I myself can readily cause to occur violate Ramsey's condition of "ethical neutrality" and illustrate this limitation. My partial belief in the event that *I will clean my cellar next week* cannot be assigned a probability via lotteries: *$1 if I clean my cellar next week, $0 else* would not be preferred to *$1 if Heads with a fair coin, $0 else*; but multiplying the rewards in both lotteries by 1,000,000 would change my preference; For $1,000,000 I will clean my cellar net week.
3. The theory describes convergence via observation, but says nothing about other ways of reaching consensus in the face of uncertainty. One can obey the rules of logic and still be "irrational" in a wider yet less well-defined sense of the term. The same holds for partial belief.

Imprecision, Probability Intervals, and Fuzziness

Imprecise probabilities were introduced by Walley (1991) as differences in certainty equivalents when buying and selling lotteries. Together with fuzzy sets, they have appeared in the climate literature (see, for example, Fu et al. 2005; Hall et al. 2007; Ghosh and Mujumdar 2009; Kriegler et al. 2009). Without disputing the substantive contribution these authors have made, I briefly discuss the use of imprecise probabilities in uncertainty accounting.

The idea is that we, or experts, cannot assess a precise degree of belief, or precise subjective probability ***P*** that:

Contribution to sea level rise from ice sheets will exceed 1 meter in 2100.

Instead, experts should give an interval, say [0.1, 0.5] in which ***P*** is certain to lie. The bounds of this interval are quite precise, but second, third, or higher order imprecisions can deal with that. Why not take a distribution over this interval and use its expected value as an estimate of ***P***? That would confuse imprecision with probability, goes the response: you can't put a probability on imprecise numbers. Denying the applicability of probability distributions within probability intervals has been delucidated as follows by Ferson et al. (2007):

The interval lacks any concentration of probability or likelihood within the interval, so the actual value is not more likely to be at any one place or another within the interval. But neither is there necessarily a uniform probability distribution over the interval. Instead the actual value has complete latitude to be anywhere within the interval with probability one.

Bounding or "simple interval measures" of course are not new; the question is the extent to which they can aid complex uncertainty accounting. The *Probabilistic Risk Assessment Procedures Guide* (US Nuclear Regulatory Commission 1983, 12) gets it right:

The simplest quantitative measure of variability in a parameter or a measurable quantity is given by an assessed range of the values the parameter or quantity can take. This measure may be adequate for certain purposes (e.g., as input to a sensitivity analysis), but in general it is not a complete representation of the analyst's knowledge or state of confidence and generally will lead to an unrealistic range of results if such measures are propagated through an analysis.

If different probability intervals are generated by different experts, what do we do with them? Hall et al. (2007) take weighted combinations of the lower bounds to be "conservative." If one has no way to distinguish between good and poor probability interval assessments, then one can do anything else with equal justice. The gates of obfuscation are thrown wide open.

Fuzziness as a representation of uncertainty has also appeared in the context of emissions scenarios (Fu et al. 2005). Regarding the question of whether fuzziness represents uncertainty, the discussion can be very brief. Suppose you get an email from an unknown person named Quincy, and you are equally uncertain whether Quincy is a man or a woman (Cooke 2003). The uncertainty that Quincy is a man would be represented by a fuzzy membership function, $\mu_M(Q)$, taking a value in the interval [0, 1] that reflects the degree to which Quincy is believed to be in the set of men (M). A similar function, $\mu_W(Q)$, describes the degree to which Quincy is believed to be in the set of women (W). The uncertainty that Quincy is a man AND a woman would be represented as the membership $\mu_{M\cap W}(Q)$ in the intersection of the sets of M and W. In the original theory, this would be the minimum of $\mu_M(Q)$ and $\mu_W(Q)$. Since you are equally uncertain whether Quincy is a man or a

woman, $\mu_M(Q) = \mu_W(Q) = ½$; and Quincy is a man AND a woman with value ½. It may well be that this sort of heuristic underlies National Research Council's reasoning when concatenating high confidence statements about climate change. Fuzzy combination rules have proliferated, but they all share this feature: the uncertainty of belonging to an intersection of two sets (M and W) is some function of the uncertainties $\mu_M(Q)$ and $\mu_W(Q)$ and does not depend on M and W themselves. Partial belief simply does not behave that way.

Deep and Knightian Uncertainty

Denying the application of probability within intervals of imprecision is related to the notion that there are "deep" uncertainties which defy quantification. One often hears that climate change is rife with deep uncertainty. The first documented use of the term appears to be the 2003 Senate testimony of the late Stephen Schneider:

"In fact, the climate change debate is characterized by deep uncertainty, which results from factors such as lack of information, disagreement about what is known or even knowable, linguistic imprecision, statistical variation, measurement error, approximation, subjective judgment, and disagreement about structural models, among others" (see Moss and Schneider 2000).

A search for a precise definition came no further than: "*By deep uncertainty we mean uncertainty that results from myriad factors both scientific and social, and consequently is difficult to accurately define and quantify"* (Kandlikar et al 2005). For an uncertainty analyst, that is standard fare. Although Moss and Schneider (2000, p. 36) advocate a Bayesian or subjectivist approach in which experts assess their subjective probability distributions, "deep uncertainty" seems to have morphed into apology for not quantifying uncertainties at all. Modelers assigning values of dubious pedigree to variables need not concern themselves with quantifying uncertainty because it is deep. The sobriquet is that deep uncertainty is "Knightian."

Economist Frank Knight's book, Risk, Uncertainty, and Profit appeared in 1921, the same year as John Maynard Keynes' Treatise on Probability, well before F.P. Ramsey's (1931) operational definition of partial belief in terms of observed preference behavior and before R. von Mises' (1928) frequentist interpretation of probability. The latter works have framed most of the subsequent discussions on the foundations of probability. Keynes believed that probabilities of various events were incommensurable and that probabilities should be organized as partial orderings[28]. For Knight, "risk proper" is measurable by resolving outcomes into equi-probable alternatives (Knight 1921, III.VII.34). By aggregating risks, the law of large numbers converts losses into fixed costs, and these would not give rise to profits. "Uncertainty," in contrast, concerns "partial knowledge" for which "the conception of an objectively measurable probability or chance is simply inapplicable" (Knight 1921, III.VII.47). Many authors have seized on such statements to argue that uncertainty in climate change is unquantifiable. A typical example is Claude Henry (2006), who also performs a post-mortem baptism of Heisenberg into his faith:

Keynes and Knight make a clear distinction between two kinds of uncertainty: the first one, called risk, may be characterized by probabilities, while this is not possible for the second one. Here we deal with decision-making under genuine uncertainty, no probability distributions being available ….Uncertainty in quantum mechanics is strictly probabilistic, and Werner Heisenberg, had he been an economist, would have called "principle of risk" his famous Uncertainty Principle.

[28] Partially ordered sets can be embedded in Boolean algebras, meaning that probabilities on a partially ordered set can be viewed is incompletely specified ordinary probability distributions.

Many economists, not the least of whom is Sir Nicholas Stern (2008), have averred that where we don't know the probability distribution, then "Knightian uncertainty" kicks in, which cannot be characterized by probabilities. Regrettably, these authors did not read further in Knight (1921, III.VIII.1): "*We can also employ the terms 'objective' and 'subjective' probability to designate the risk and uncertainty respectively, as these expressions are already in general use with a signification akin to that proposed.*" Knight, writing in 1921, did not know how to measure subjective probabilities. Neither did he know how to measure "risk" or objective probabilities. The Laplace (1814) interpretation of probability to which Knight's definition of risk appeals, has been moribund, if not dead, since the end of the 19th century. It was, after all, von Mises (1928) who emphasized that objective probabilities can be measured as limiting relative frequencies of outcomes *in a random sequence*. Like many authors of this period, Knight appears to have been unaware of the role of (in)dependence assumptions and believed that objective probabilities are much more objective than modern probability warrants. It is indeed significant that economists claiming that climate uncertainty cannot be described with probability harken back to a period when probability, both objective and subjective, was not well understood.

Nonetheless, Knight was ahead of his time. The idea of calibrating the probability judgments of the individual "business man" prefigures the modern use of structured expert judgment (Knight 1921, III.VII.43):

"A still more interesting complication, and one of much greater practical significance, is the possibility of forming a class of similar instances on entirely different grounds. That is, instead of taking the decisions of other men in situations more or less similar objectively, we may take decisions of the same man in all sorts of situations. It is indisputable that this procedure is followed in fact to a very large extent and that an astounding number of decisions actually rest upon such a probability judgment…."

3.2 Shallow Uncertainty

We have witnessed the emergence of deep or Knightian uncertainty, imprecise uncertainty, fuzzy uncertainty and all the others mentioned in the previous chapter. Wicked uncertainty is also out there but not represented in the *Proceedings of Uncertainty in Artificial Intelligence*. Surely there is more awaiting discovery. In the spirit of ecumenism I hazard a short list:

Black Hole Uncertainty - uncertain about your uncertainty about your uncertainty...to the power of the largest infinite cardinal number, which is uncertain – once sucked in, no escape.
Mobius uncertainty - when both sides of an issue are the same
Russellian uncertainty - when each of two contradictory statements are true
Rumsveldian uncertainty - when you don't know whether you know or not
Teleological uncertainty – uncertainty about what's the point?
Zenoian uncertainty - learn as much as you like and remain equally uncertain
Scatological uncertainty - makes you soil your drawers.

Surprisingly, almost no attention has been given to shallow uncertainty. Is that the uncertainty described in elementary textbooks in probability and statistics, where one learns to reason probabilistically? Could that be the reason the uncertainty sleuths pass it over? Perish the thought. I propose the following definition: *Shallow uncertainty is uncertainty resulting from undefined terms, careless formulation, lack of operational meaning and overall intellectual sloth.* The good news is that shallow uncertainty, originating from our own insouciance regarding the meaning of words, is much easier to remove than uncertainty about our future climate.

The Social Discount Rate tells how future damages should be discounted back to the present. It is recognized as an important driver, if not *the* most important driver, in the economic models for climate change. SDR is often written as:

$$SDR = \rho + \eta G(t),$$

where ρ is the rate of pure time preference, η is the coefficient of constant relative risk aversion, and G(t) is the time average growth rate of per capita consumption out to time t. Some (e.g., Stern 2008) see a strong normative component. Others infer values for ρ and η from data (Evans and Sezer 2005). Nordhaus (2008) equates the SDR to the observed real rate of return on capital with a constant value for G(t) and sees ρ and η as "unobserved normative parameters" (Nordhaus 2008, p. 60) or "taste variables" (p. 215) that are excluded from uncertainty quantification. Pizer (1999) assigns distributions to ρ and η. Nordhaus and Popp (1996) put a distribution on ρ. Weitzman (2001) fits a gamma distribution to SDR based on an expert survey. Frederick et al. (2002, 352) note that "virtually every assumption underlying the DU [discounted utility] model has been tested and found to be descriptively invalid in at least some situations." They also cite the founder of DU, Paul Samuelson: "It is completely arbitrary to assume that the individual behaves so as to maximize an integral of the form envisaged in [the DU model]" (Frederick et al. 2002, 355). Weitzman (2001) and Newell and Pizer (2003) show that uncertainty in the discount rate drives long-term rates down. We may distinguish variables according to whether their values represent

a) Policy choices
b) Social preferences
c) Unknown states of the physical world.

Uncertainty quantification is appropriate for (b) and (c) once operational definitions are supplied; it is not appropriate for (a) if we are the choosers of policies.

What is the operational meaning of society's rate of pure time preference. That is, how, with sufficient means and license, would this be measured? Similar questions pose themselves regarding society's coefficient of risk aversion, the utility of consumption, the preferences of a representative consumer, etc. If the modeling community has not agreed on operational definitions for such terms, then the models are not ready for uncertainty analysis. Quantifying and propagating uncertainty in scientific modelling requires a great deal of effort in clarifying the operational meaning of the uncertain terms. If we don't know what it means, we can't be uncertain about it. But we can still obfuscate.

Good News

The discount rate is closely bound up with the rate of economic growth. Good news, the shallow uncertainty about the growth rate has been removed! Structured expert judgment has been applied to quantify uncertainty for several modeling parameters out to 2300 (Rennert et al 2022) and is getting hit hard in the clickosphere. Two panels of economist experts were convened at Resources for the Future and gave probabilistic assessments in the growth of GDP per person and CO2 concentrations out to 2300. The statisical accuracy of the panelists was among the best in the annals of structured expert judgment.

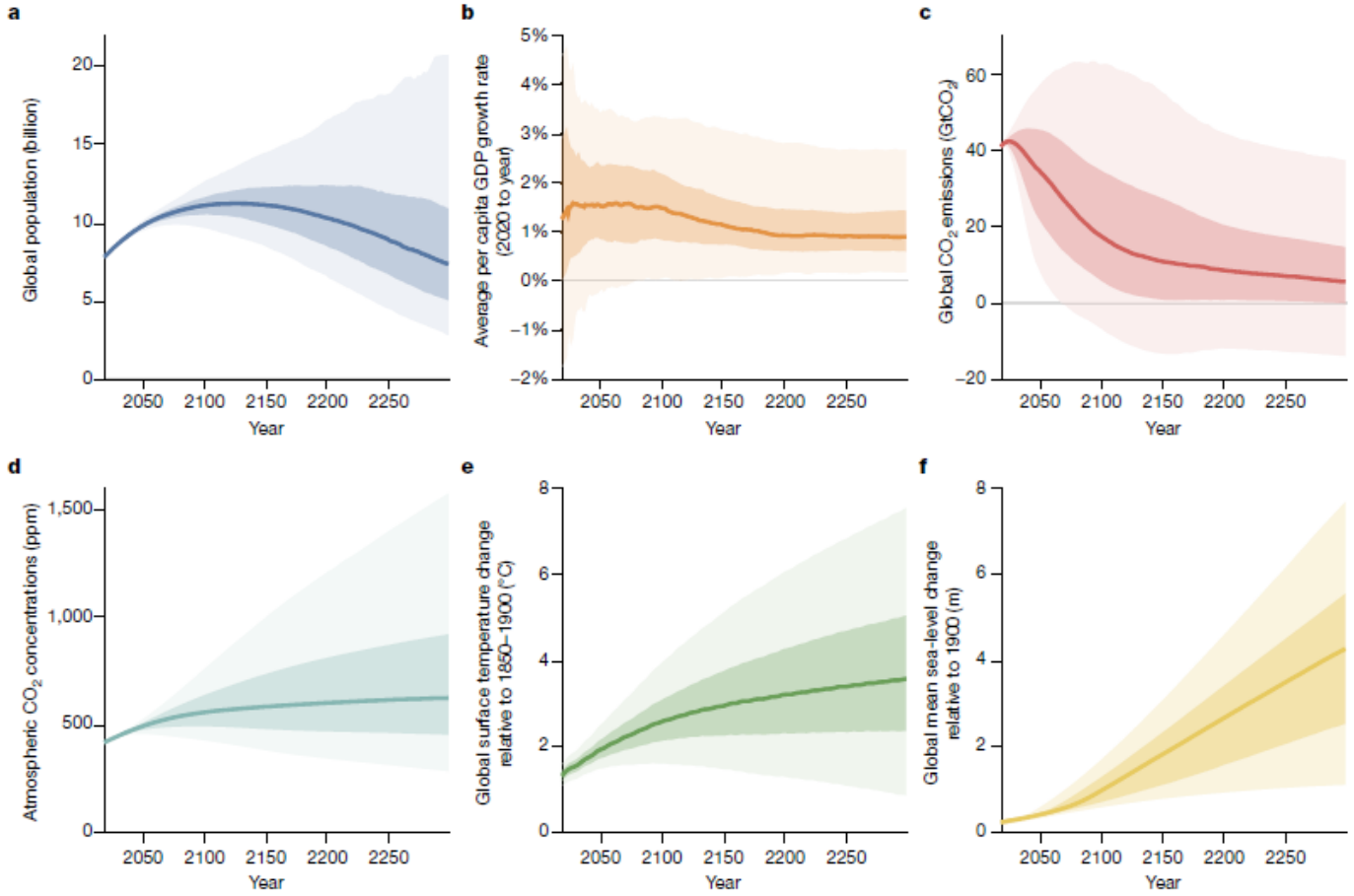

Fig. 1 | RFF-SP socioeconomic scenarios and the resulting climate system projections. a–c, Probabilistic socioeconomic projections for global population (**a**), per capita GDP growth rates (**b**), and carbon dioxide emission levels (**c**) from the RFF-SP scenarios. **d–f**, Corresponding climate system projections that account for parametric uncertainty in FaIR and BRICK for atmospheric carbon dioxide concentrations (**d**), global surface temperature changes relative to the 1850–1900 mean (**e**), and global mean sea-level changes relative to 1900 (**f**). In all panels, solid centre lines depict the median outcome, with darker shading spanning the 25%–75% quantile range and lighter shading spanning the 5%–95% quantile range.

Economists, for whatever reason, are on the whole good at probabilistic reasoning. Details of the elicitation are found in Rennert et al 2021), Figure 2 summarizes performance.

Expert Scores EGS					
expert	Statistical accuracy	mean information	# variables	weight	Rel.Inf to EW DM
1	0.706	0.673	11	0.3	0.433
2	0.399	0.829	11	0.209	0.608
3	0.008	0.894	11	-	0.6
4	0.197	0.659	11	0.082	0.363
5	0.215	1.094	11	0.148	0.711
6	0.327	1.131	11	0.233	0.761
7	0.154	0.291	11	0.028	0.331
8	0.018	1.087	11	-	0.671
9	0.0003	0.727	11	-	0.511
PW05	0.492	0.457	11	-	
EW	0.37	0.266	11	-	

Figure 2: Scores and weights for all 9 EGS experts when performance weights are not optimized but computed for the six weighted experts with statistical accuracy > 0.05. PW05 denotes the performance weighted combination combining all experts with statistical accuracy greater than 0.05 and weights proportional to the product of statisticcal accuracy and mean information, in accord with the theory of proper scoring rules. EW denote the equal weight combination (Rennert et al 2021).

The picture for the 10 experts in the Future Emission Scenarios is similar.

3.3 Scientists Sentenced to 6 Years in Prison for "Optimizing" Uncertainty of L'Aquila Earthquake[29]

It is a truism in risk management that every disaster was predicted by someone, sometime, somehow. The April 6, 2009 earthquake that devastated the Italian city of L'Aquila was "predicted" (actually retrodicted: evidence was adduced post hoc) by anomalous toad behavior 70 km away. An earthquake in the Abruzzo region was also predicted by heretic scientist Giampaolo Giuliani based on elevated concentrations of radon gas. Guiliani was cited for *procurato allarme*—essentially instigating public alarm or panic — and forbidden from making any public pronouncements just prior to the L'Aquila quake.

Initially ostracized, Giuliani has since been invited to give talks by the American Geophysical Union. His rehabilitation is rapid by Italian standards; Galileo waited over 350 years. Another celebrated prognosticator, Raffaele Bendandi, nailed a few earthquakes, but was less fortunate with his 1959 discovery of a new planet between Mercury and the Sun which he named after his home town Faenza.

If predictions are not based on sound science, then the prognosticator's forecast must be judged in the context of ALL his/her other forecasts. Radon prediction is currently promoted by physics Nobel laureate Georges Charpak, but has yet to demonstrate predictive validity for earthquakes. As *Nature* has reported:

The recent ICEF (International Commission on Earthquake Forecasting) report deemed Giuliani's findings 'unsatisfactory', and he has yet to publish a single peer-reviewed paper on his radon work. Nonetheless, he maintained an open website that posted real-time radon measurements from his detectors, and in interviews with journalists and in an informal mobile-phone network, Giuliani made predictions about low-level seismic activity. Although the ICEF report notes that he made two false forecasts, The Guardian newspaper dubbed him 'The Man Who Predicted An Earthquake', after the April 2009 quake hit.

Seven members of the "Major Risks Committee", one of whom reassured the public of the unlikelihood of an earthquake shortly before of the April 6 shock, were found guilty Monday of multiple manslaughter, sentenced to 6 years in prison and fined a total of 7.8 million euros (about 10 million dollars). The public prosecutor, Fabio Picuti, argued that although the committee members could not have predicted the earthquake, they had translated their scientific uncertainty into an overly optimistic message.

The verdict will surely be appealed but it has already heightened interest in Structured Expert Judgment (a method for synthesizing and communicating scientific uncertainty). A testimonial by volcanologist Willy Aspinall us summarized in a Nature Communication.

When politicians and other decisionmakers seek advice from experts, they hope to receive something unambiguous that they can act on. But . . . in complex situations—assessing the risk of volcanic eruptions and earthquakes for instance—attempts to force a consensus can be counterproductive. It is better to quantify the uncertainty that exists and build it into the decisionmaking process. The way to do this, says Aspinall, is through "expert elicitation," specifically by using a method which weighs the opinion of each expert based on his or her knowledge and ability to judge relevant uncertainties.

[29] Based on the authors RFF blog of Oct.24, 2012 https://www.resources.org/common-resources/scientists-sentenced-to-6-years-in-prison-for-quotoptimizingquot-uncertainty-of-laquila-earthquake/

Garbling, misrepresenting or even suppressing uncertainty is a recurring feature of failed communication between science, decision makers and the public. As Aspinall notes, of structured expert judgment is designed to "quantify uncertainty, not to remove it from the decision process".

The verdict could also divert attention from the shoddy building construction responsible for many of the 309 deaths of the L'Aquila earthquake.

4. Science Under Uncertainty

If the science "isn't there yet," scientists are supposed to disagree—without disagreement, science could not advance and we would still be computing epicycles. Confronted with unwelcome scientific advice, interested parties may seek out, or in some cases even generate, conflicting scientific views to neutralize the unwelcome impact (Oreskes and Conway 2010). Lacking the ability to evaluate the advice, public media striving for balance can unwittingly promote the idea that conflicting advices can both be ignored. Certain parties can use this situation to marginalize scientific input altogether.

"Nearing a tipping point? No, no. Some say that and some say differently. I mean, you have scientists on both sides of it. …. I have a natural instinct for science, " **Trump** Oct 17 2018

The resulting picture of scientific debate may be termed destructive disagreement in which science is reduced to conflict over resources.

Figure 1: Destructive disagreement over the cause of the pause in global warming

When science is marginalized, climate policy is left in the hands of leaders like US Senator (1994 - 2023) James Inhofe who chaired the U.S. Senate Committee on Environment and Public Works (EPW) from 2003 to 2007 and again from 2015 to 2017.

"We should have no concern for the environment because after the great flood with Noah, god promised that he would never ruin the earth again" **Sen. James Inhofe**

The 19th century scientific method promises rational conflict resolution, but only after the facts are in. With destructive disagreement, that may take more time than we are alloted. Probabilistic reasoning enables the pooling of information, reduction of uncertainty and the probabilistic resolution of scientific conflicts. As every investor and every decision analyst knows, uncertainty costs money and cost-free uncertainty reduction is always worthwhile. It may not always be worth the costs, but we have to compute the costs and benefits to know.

4.1 Conflicting Measurements Reduce Uncertainty in Climate Sensitivity[30]

This chapter probabilistic conflict resolution involving Equilibrium Climate Sensitivity (ECS), made possible by groundwork laid by researchers a NASA. The economic benefit of reducing uncertainty in ECS have also been quantified (Cooke et al 2013, 2015, 2016) and runs into Trillions. All measurements have error, in this case the error distributions have been quantified. The scatter in a telescope can be measured by studying the telescope and the situation in which it is used. When we look at a distant planet through a telescope, we don't see the planet's position, we see the distribution of the planet's position caused by interaction between photons reflected by the planet and the measuring instrument. We learn about that instrument error by studying our instruments (recall Kant chap. 1). For the mathematicians this is accomplished by Renyi conditionalization (Cooke and Wielicki 2018, Renyi, 1970). Different instuments have different errors and the errors can be positively or negatively correlated. Error distributions can be combined probabilistically. They contain information which we ignore at our peril.

Measuring Equilibrium Climate Sensitivity
Equilibrium Climate Sensitivity (ECS) is the amount by which the Earth's mean surface temperature will eventually rise upon doubling the atmospheric concentration of CO2. We want to design measurements to improve on the initial Roe Baker uncertainty distribution adopted by the US Government's computing platform for the Social Cost of Carbon (IWGSCC 2009, 2013, 2016). The computations use a Belief Network, or Bayesian Belief Network to incorporate future measured values and update the initial uncertainty distribution. These future values have of course not yet been measured; we use hypothetical future measured values to illustrate probabilistic conflict resolution.*(A video provides a demonstration of the software used for these computations, which is freely available at http://www.lighttwist.net/wp/).*

Most of what we measure are proxies for the real quantities of interest. For past global temperatures we measure tree rings, ice cores, coral bands, isotope fractionation, and so on. ECS is our variable of interest. The 2016 interagency working group memo on the social cost of carbon proposed an initial distribution for ECS with expected value of ECS at 3.3°C. Given a business-as-usual emissions scenario, one of the proxies is the Decadal global mean surface Temperature Rise (DTR). A linear increase of 0.2°C per decade would lead to 2°C warming in 100 years. Imagine how difficult it is to measure a 0.2°C change in average surface temperature over 10 years. Another proxy is the decadal percentage change in cloud radiative forcing (CRF), the ability of clouds to alter the Earth's reflected solar energy or thermal radiation emission to space, thereby affecting the amount of global warming by affecting the Earth's energy balance. NASA researchers have analyzed the error in the existing and proposed new enhanced systems with the goal of demonstrating how these systems could, over time, reduce our uncertainty about ECS. Recall, uncertainty costs money. Reducing uncertainty can save money.

Satellite weather data is used in this example, detailed in (Cooke and Wielicki 2018). The IPCC does not use satellite data for its surface temperature trends relative to ECS. Instead it uses multiple lines of evidence, surface air temperature observations, ship observations, and weather balloon observations. These are not calibrated extremely well. The IPCC relies on the fact that many different observing systems give similar results. Without rigorous error quantification it is impossible to pool this information; these remain separate lines of evidence and uncertainty regarding ECS is not reduced

[30] Based on RFF blog by Roger Cooke and Bruce Wielicki https://www.resources.org/common-resources/conflicting-measurements-reduce-uncertainty-in-climate-science-we-can-work-it-out/ and on Cooke, R.M. & Wielicki, B. (2018) Probabilistic reasoning about measurements of equilibrium climate sensitivity: combining disparate lines of evidence, Climatic Change, 151(3), 541-554 https://doi.org/10.1007/s10584-018-2315-y

Figure 2 assumes a launch of the satellites in 2020 and uses information collected out to 2030. According to the climate models of the IWGSSC, if we knew the value of ECS exactly we could predict DTR and CRF with certainty. However, we must go the other way, we observe DTR and CRF for 10 years taking account of their uncertainties and adapt our initial uncertainty about ECS. This is called conditionalizing or updating the initial distribution of ECS on the measured values. Ten years isn't very long for observing decadal trends. The natural variability in the Earth's climate system will cause these measured values to deviate from the long term trends. The effect of extending the observation times can be calculated based on yearly fluctuations. The numbers shown on the corresponding arrows are the correlations between the distribution of ECS and the measurement results in 2030. Natural variability in the climate system causes these correlations to be weak, but they become stronger as we observe longer. On the bottom row are the actual measurement satellites currently in operation (green) and new enhanced measurement satellites designed by NASA (yellow).

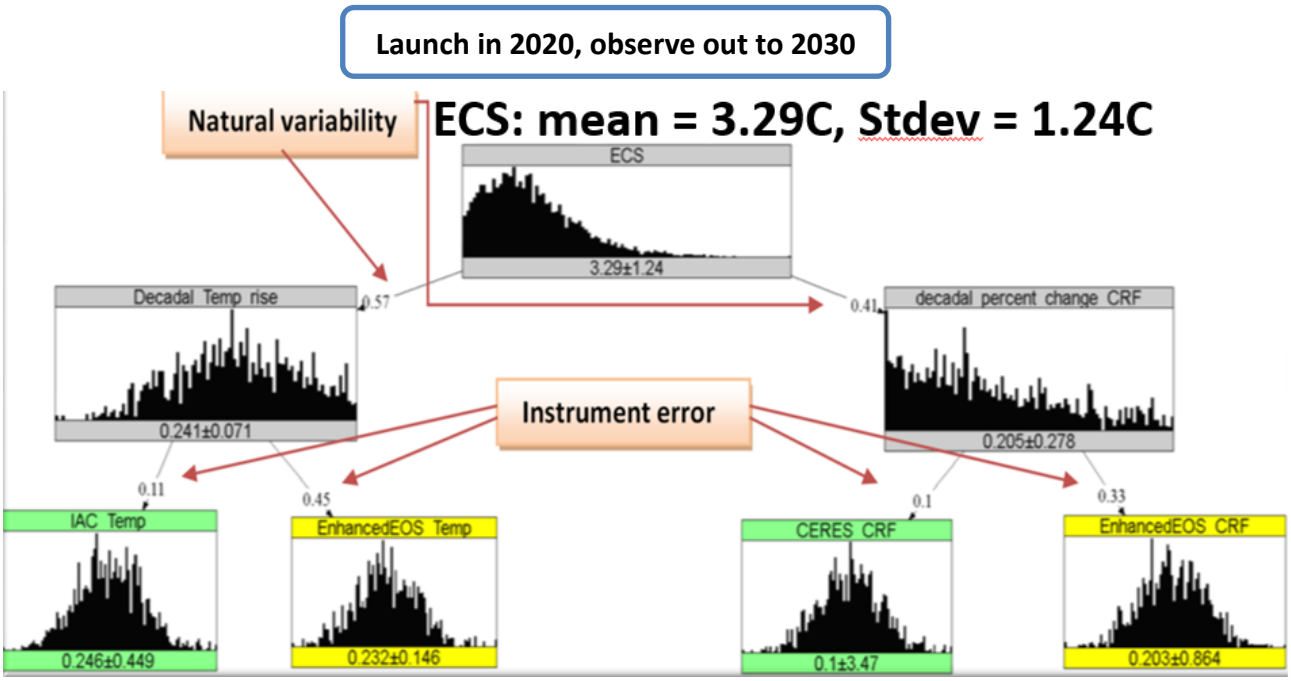

Figure 2: Correlations in 2030 following a launch in 2020. DTR and decadal percentage rise in CRF are deterministic functions of ECS for a given emissions scenario and Business as usual.

These satellite systems are also subject to instrument error caused by things like limited orbital sampling and calibration drift. These errors are quite large after 10 years resulting in very weak correlations between what we actually measure (bottom row) and what we would see if there were no instrument error (middle row). These errors can also be attenuated with longer observation times. Figure 3 shows the situation after 30 years of observations. All correlations have gotten stronger.

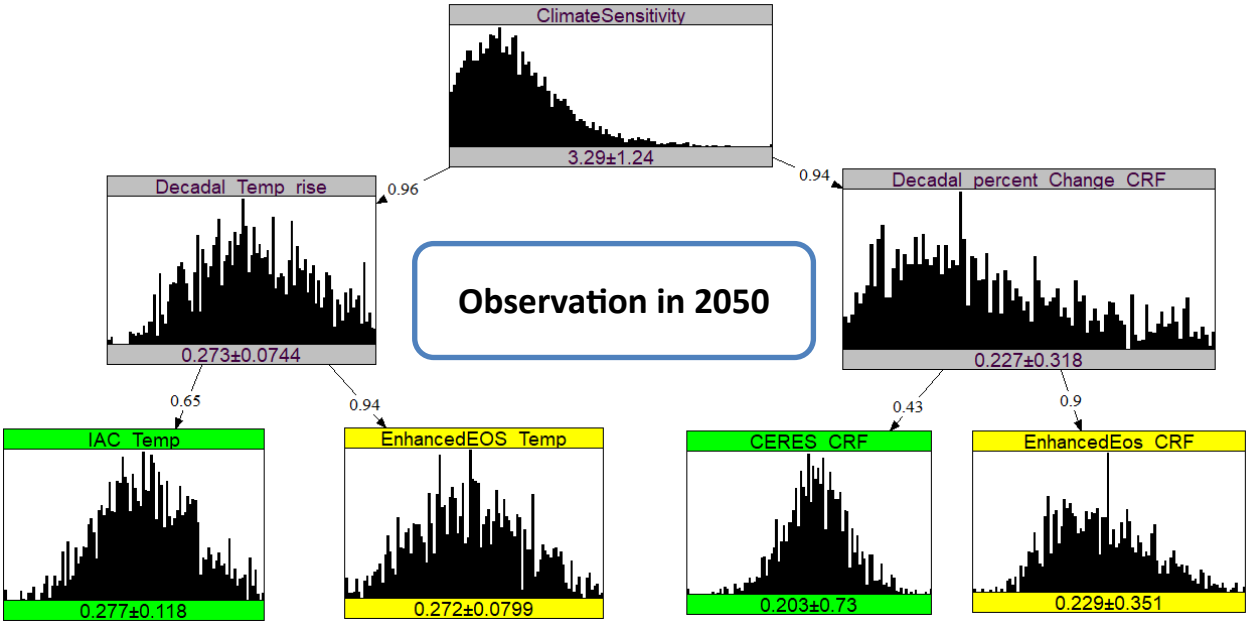

Figure 3; Observation in 2050 after launch in 2020.

Returning to 2030, Figure 4 shows the enhanced system for DTR returning a very large and noisy value of 0.524 in 2030, corresponding to a very high increase in 100 years of 5.24ºC. It is large but based on these distributions certainly not impossible. Propagating this measured value through the belief network produces a shift in the distribution of ECS relative to the original distribution (gray); updated distribution for ECS with mean 4.01C and standard deviation 1.46C. It also causes a small shift in the distribution of decadal percent change in CRF; the mean value of 0.205 has become 0.27 and the standard deviation 0.27 has become o.29 (the original gray distribution is largely masked). The green and yellow distributions for CRF have also shifted slightly.

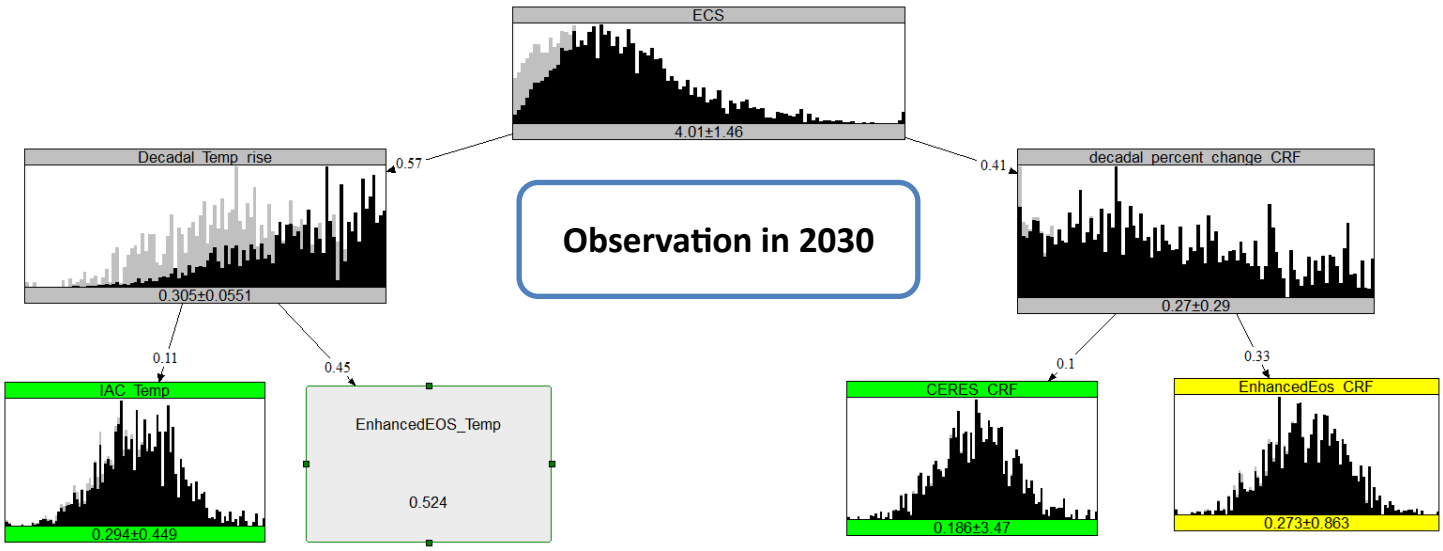

Figure 4: Figure 2 is conditionalized on a high measured value for DTR in the enhanced measuring system in 2030.

The uncertainty in ECS as reflected in the standard deviation increased from 1.24C to 1.46C. This is quite impossible within the simple classical statistical error analysis, but is not uncommon and is termed *negative*

learning[31] (Oppenheimer et al 2008): We learn that we didn't know a much as we thought we did. The new measurement pushes the distribution of ECS up but because the measurement is very noisy it drags only part of the initial distribution with it, causing the distribution to be more smeared out as reflected in the higher standard deviation. It is weird and counter intuitive, but perfectly correct.

Figure 5 shows that existing and enhanced systems can return the same measured results in 2050 and nonetheless influence our knowledge of ECS distribution very differently. It is paradoxical only if you forget to take the noise into account. Since the enhanced systems have much smaller noise they will have greater effect on the initial distribution. The enhanced systems now give strongly discordant predictions of ECS. A malicious pundit would say that the scientists don't know what they're talking about. Nonetheless, the uncertainty accounting on which this is based is solid.

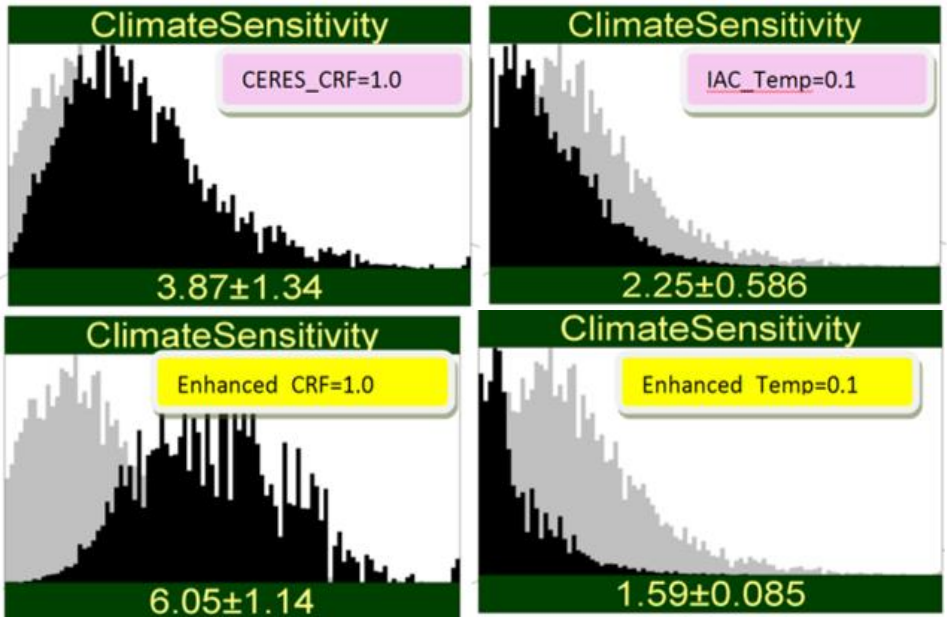

Figure 5: Existing systems (pink) and enhanced systems (yellow) return the same highly discordant measured values in 2050 but the effect is much greater in the enhanced system due to the much smaller noise.

Figure 6 shows the result of pooling the two yellow pieces of information. Look what happens: the mean value of ECS is 2.31C, close to its initial value of 3.29C, but the uncertainty has shrunk dramatically. The standard deviation has dropped from 1.24C to 0.289C. Uncertainty costs money, reducing uncertainty pays off. Disagreement pays off, marginalizing science does not. Without uncertainty accounting we would be left with the DTR team and the CRF team trashing each other in a competition for resources. This example is unusual only in that NASA researchers have already done the heavy lifting. Showing that these things are possible will hopefully motivate others to do some heavy lifting on other problems.

[31] If two measured values x1, x2 have independent error variances V1, V2, then the maximum likelihood esimate is ½(x1+x2) whose variance is ¼(V1+V2).

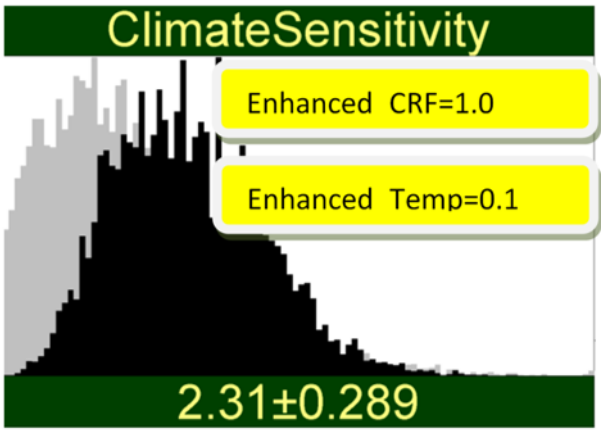

Figure 6: Conditionalizing both DTR and CRF values in 2050

4.2 Respectability Looming for Expert Judgment[32]

Expert judgment, like the Internet, runs from sublime to sordid. Since its halting entry into the Halls of Science with the Delphi studies of the 1960s, expert judgment has remained something of an embarrassment: scientists use it all the time but would rather not talk about it. Now it is poised to become respectable. Here's some leading indicators:

Bamber and Aspinall's 2013 paper "An expert judgment assessment of future sea level rise from the ice sheets" was selected in 2016 as one of ten articles to highlight research in *Nature Climate Change* over the previous five years. Healthy suspicion within the science community was allayed by the "classical model for structured expert judgment," the hallmark of which is empirical validation with performance-based weighted combinations of experts' judgments. Exhortations to take on climate uncertainty have appeared in Nature Climate Change. The National Academy of Science plea for its use in quantifying the social cost of carbon has recently borne fruit in "Comprehensive Evidence Implies a Higher Social Cost of CO2".

Other recent events also signal expert judgment's ascendency. Climate gadfly Judith Curry penned an excellent blog post in 2015 on expert judgment and rational consensus, emphasizing the risks of confusing consensus with certainty. Also in 2015, Australian biologist and bio-security expert Mark Burgman's *Trusting Judgment* hit the bookshelves, with exhaustive reviews of the sordid side of expert judgment. This followed Sutherland and Burgman's piece in *Nature* on using experts wisely and Aspinall's appeal for a "route to more tractable expert advice."

Building on the pioneering work of Eggstaff et al. on cross-validation, out-of-sample-validity was demonstrated in 2017 for the set of professional studies conducted between 2006 and 2015. Persistence of expert performance was established in 2021 (chapt 4.3). Elsewhere, highly visible applications of expert judgment appearing in top-tier scientific journals have targeted the Asian carp invasion of Lake Erie and nitrogen runoff in the Chesapeake Bay, both with out-of-sample validation. In 2016, the World Health Organization (WHO) completed a structured expert judgment study of food-borne diseases with empirical validation on industrial scale: 74 experts distributed over 134 panels averaging 10 experts each quantified uncertainty in transmission rates of pathogens through food pathways for different regions of the world. The Center for Disease Control also conducted a large study on foodborne disease pathways. A study on the effect of breast feeding on IQ was completed. A US EU joint effort in 2019 quantified ice sheet's contribution to sea level rise (Chapter 4.4) and the previously noted study on CO2 emissions and economic growth (2022).

The world of expert judgment divides into two hemispheres. The science/engineering hemisphere usually works with small numbers (on the order of 10) of carefully vetted experts, asks them about uncertain quantities with a continuous range, and propagates the results through numerical models. The psychology hemisphere estimates probabilities of future newsworthy events. Philip Tetlock's Good Judgment Project was proclaimed the winner of a five-year forecasting tournament organized by the Intelligence Advanced Research Projects Activity using the Brier Score for evaluating forecasters (disparaged in the classical model for confounding statistical accuracy and informativeness). Drawing from an expert pool of more than 3,000 and skimming off the top 2 percent of all experts, Tetlock's group distilled a small group of "superforecasters." With a small fraction of Tetlock's resources, Burgman's "Australian Delphi" method (based on the classical model with Delphi-like add-ons) is said to make a strong showing—though data and analysis from the tournament are not released.

[32] Based on the author's RFF blog of 2017 https://www.resources.org/common-resources/respectability-looming-for-expert-judgment/

In applications of the classical model, experts are typically asked to assess 5-, 50- and 95- percentiles for continuous quantities of interest—and for calibration variables (order 10) from their field, the true values of which are known post hoc. Experts are scored on statistical accuracy and informativeness. If only 2 of 10 values of calibration variables fall within an expert's 90 percent central confidence band, that would result in a low statistical accuracy score. Informativeness roughly corresponds with the degree to which an expert's percentiles are close together. (Proper definitions and data are readily available.) The two scores are negatively correlated, though the WHO data in the Figure 1 show that the correlation attenuates as we down-select to statistically more accurate experts.

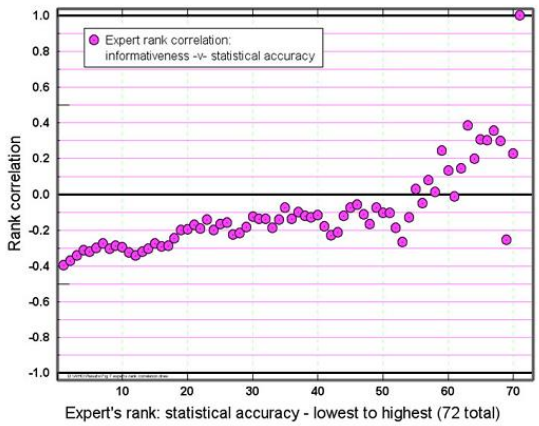

Figure 1. Rolling Rank Correlations of Informativeness and Statistical Accuracy for Subsets of Successively More Statistically Accurate Experts **Source*: Aspinall et al. (2016)***

Unlike current events, science/engineering studies do not have access to thousands of experts and years of data per expert panel. Rather, "in-sample" validation looks at performance on the calibration variables, and "cross-validation" initializes the weighting model on subsets of calibration variables and gauges performance on the complementary set. Eggstaff and colleagues developed cross-validation for the extensive database of applications with the classical model in 2014. The performance ratios for performance-based and equal weighting in Figure 2 speak for themselves. Both hemispheres agree that measuring expert performance and using performance-based combinations pay off.

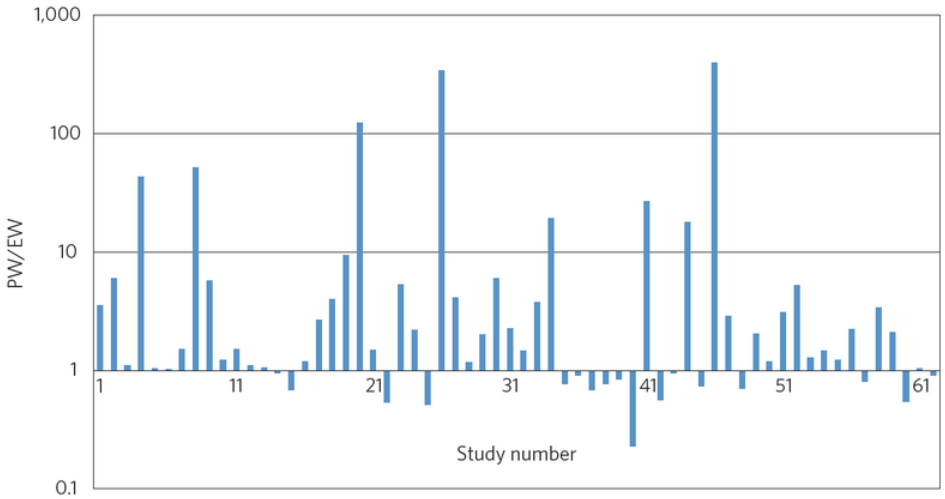

Figure 2. Performance Weight/Equal Weight (PW/EW) Ratios for 62 Studies

Note: The ratios concern combined scores for statistical accuracy and informativeness, aggregated over all test/training sets within each study. *Source*: Cooke (2015).

**At a 2013 RFF event, Bamber, Aspinall, and other experts discussed the ice sheets covering Antarctica and Greenland, which pose both the largest risk and uncertainty for sea-level rise and are considered to be one of the greatest hazards from future climate change. Watch the video.*

4.3 The Science of Forecasting: the Special Sauce[33]

"Where's this special sauce?" CIA operative Bernadette in *The Report* asks psychologists James Mitchell and Bruce Jessen, after 60 days of waterboarding Muhammad Rahim yielded no results. "You have to make this work. It's only legal if it works."

Oracles, augurs, prophets, and pundits studied entrails, birds, and dreams looking for the special sauce, but the science of forecasting is young. A new article in the *International Journal of Forecasting* claims to have found the special sauce—or at least some key ingredients—in the classical model for structured expert judgment. Torture isn't among them.

Good expert forecasting does not correlate with citations, status, or blue ribbons. If you want to pick a good prophet, you look at her record. But her record of what? Tallying previous forecast errors isn't as simple as it sounds. An error of three years in guessing the age of a Boeing-737 is not the same as an error of three years in guessing the age of your youngest daughter. Selecting forecasters based on previous errors requires converting errors to a common scale that accounts for the difference between aircraft years and daughter years.

Here's the first key ingredient of the special sauce: *the common scale is probability*. How likely is an error of three aircraft years or three daughter years? That would be easy if we had a universal likelihood scale for all responses, but alas, we don't. The next-best thing is the subjective probability of the expert forecaster herself. Suppose that for every forecast of a quasi-continuous unknown quantity, we tally how often the true value, revealed after the fact, falls within the expert's 90 percent confidence bands.

Why would this help? Aren't all experts pretty much equally able to assess how much they know? As it turns out, no. For some experts, saying they're 90 percent certain that the true value falls in some interval confers a less than 10 percent chance of that predicted result actually happening. For other experts, you can take that 90 percent to the bank. Understanding an expert's ability to gauge the likelihood of her own predictions might be good to know before we bet the farm on her advice.

Equipped with a common scale, we can reveal the second key ingredient of the special sauce: *experts are not equal in their ability to quantify uncertainty.* But how do we measure "ability to quantify uncertainty," and how we can use that information?

The third and perhaps most surprising sauce ingredient can help us: *it's better to combine expert uncertainties than to combine their point forecasts, and it's better still to combine expert uncertainties based on their past performance.*

Example: Center for Disease Control
In a recent expert elicitation on foodborne illness pathways at the Centers for Disease Control[34] and Prevention, one of the 14 calibration questions is the following: "Between January 1, 2016, and December 31,

[33] Based on Cooke, Roger M., Marti, Deniz and Mazzuchi, Thomas A., (2021) Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say? International Journal of Forecasting, published online July 25, 2020, https://doi.org/10.1016/j.ijforecast.2020.06.007

[34] Based on Beshearse, Elizabeth, Beau B. Bruce, Gabriela F. Nane, Roger M. Cooke, Willy Aspinall, Tine Hald, Stacy M. Crim, Patricia M. Griffin, Kathleen E. Fullerton, Sarah A. Collier, Katharine M. Benedict, Michael J. Beach, Aron J. Hall, Arie H. Havelaar (2021) "Attribution of Illnesses Transmitted by Food and Water to Comprehensive Transmission Pathways Using Structured Expert Judgment, United States" , Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 27, No. 1, January 2021, DOI: https://doi.org/10.3201/eid2701.200316

2016, a total of 11,277 samples of raw ground beef from 1,193 establishments were tested for *Salmonella spp*. Of these samples, how many tested positive for *Salmonella spp*?" The medians and the 5th and 95th percentiles of the 48 participating experts are shown in Figure 1, as is the true value: 200 samples tested positive. Had we simply asked the experts for their best guesses, we would have gotten something like the average of the 48 median values, which is 553.

If we also ask for their 90 percent confidence bands, then we have the option of combining their distributions. If we weight the experts equally (equal weighting; EW) and take the median of this combined distribution, we get 286 predicted positive samples, with a 90 percent confidence band that ranges widely, between 1 and 4,769. We can also weigh the experts based on their performance on all calibration variables (performance-based weighting; PW). The median of the PW combination is 230, with a 90 percent confidence band of 60–350.

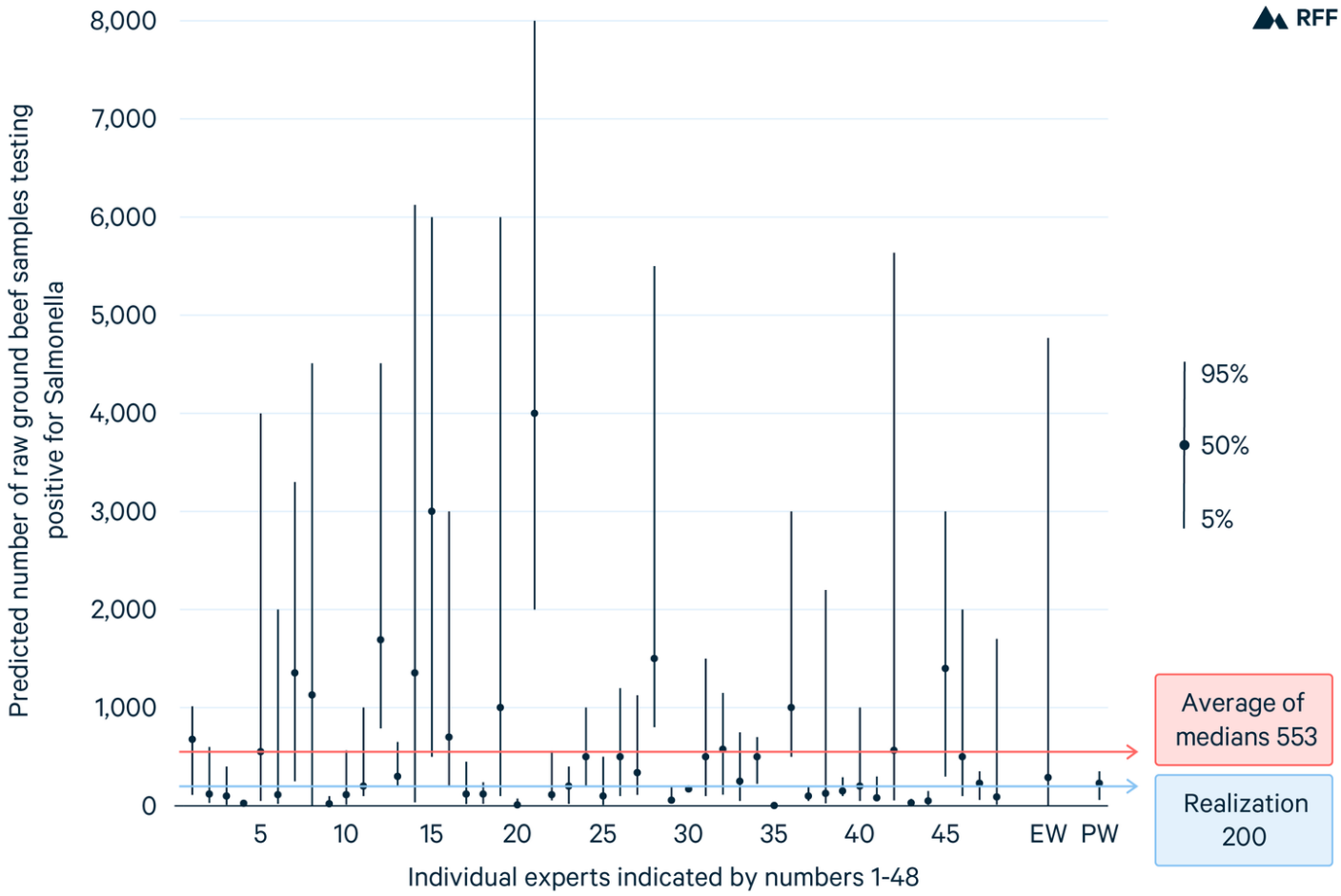

Figure 1 Predictions from 48 experts in a recent elicitation on foodborne illness pathways by the Centers for Disease Control and Prevention. Depicted is each expert's prediction, along with the 90 percent confidence bands for their predictions, as reported by each expert. Shown are the medians and average of the medians across experts (553) for equal weighting (EW) combinations, along with the true value (200) and median prediction of performance-based weighting (PW) combinations

Of course, this is just one example chosen to illustrate these features; we need more examples to draw accurate conclusions. Our article chronicles several applications of the classical model. Other highlighted examples of

professional applications include nuclear safety with the European Union and the United States Nuclear Regulatory Commission; fine particulates with Harvard University and the government of Kuwait in 2004–2005; foodborne diseases for the World Health Organization in 2011–2013; ice sheet dynamics; and volcanic hazard, forecasts, and risk level assessments in different parts of the world. From 2006 to 2018, 49 commissioned applications have involved 530 experts, who are scored by assessing 580 calibration variables from their fields. The judgments of about 75 percent of these experts, considered as statistical hypotheses, would be rejected at the traditional 5 percent significance level.

What have we Learned?
So, what have we learned? Is one expert as good as another? Are the measured performance differences just noise? This sounds like the beginning of a fruitless debate, but we can actually test the ideas with performance data.

Notice that performance-blind combination schemes, such as equal weighting, are unaffected if the assessments in an expert panel are randomly reallocated item-wise among the experts. Performance weighting depends on being able to identify high-performing experts. If we randomly reallocate the assessments, then any remaining performance differences will be due to "noise". In contrast to persistent influences like knowledge, experience, and intuition, random stressors might be things like fatigue, mood, distraction. It is not possible to observe or measure these influences. We can however test the claim that differences are just "noise". The null hypothesis is that the experts' responses for each variable are independently sampled from the same distribution (over distributions). If the experts were re-elicited a short time later, then their responses would be independent re-samples from this distribution. This is the operational meaning of the statement that expert differences are not persistent.

For each of the 49 studies, we repeat this random scrambling 1,000 times. How does the statistical accuracy of the best-performing expert in the original panel compare with the best performers in each of the 1,000 scrambled panels?

If there really were no difference between the real and scrambled best performers, then the real best performer could just as well be any of the 1,000 scrambled best performers: as such, (s)he has a 50 percent chance of performing better than the median of the scrambled best performers, a 30 percent chance of outperforming 70 percent of the scrambled best performers, and so on. Among the 49 studies, we would expect the best performer to have better statistical accuracy than 50 percent of the other experts in 24.5 (i.e., 50 percent) of the studies. If we look at the data in Figure 2, we readily see that this is not the case. In 37 studies, the original best performer has higher statistical accuracy than one half of the scrambled best performers. If the origifnal and the scrambled panels were really the same, the chance of seeing a distribution at least as lopsided as that in Figure 2 is 0.00024. Other performance metrics, in particular having the lowest absolute percentage error, give a similar picture. More powerful statistical tests bring the probability down to E-12 of seeing this result caused by noise. The performance differences among experts are real.

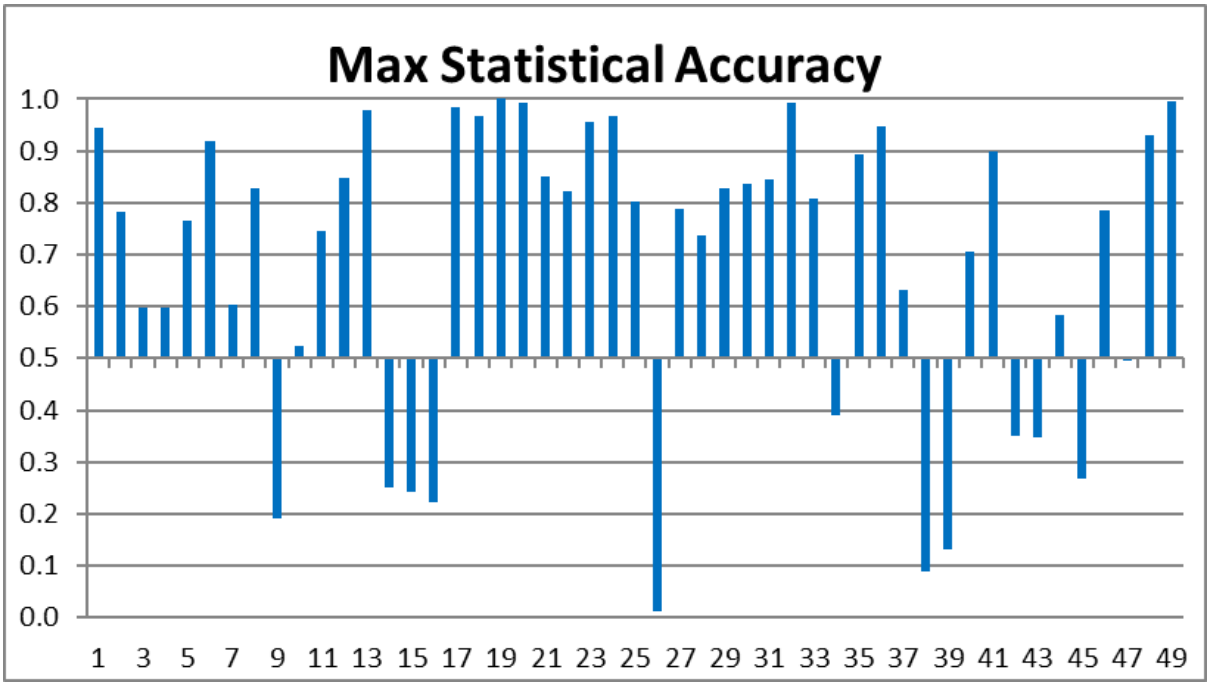

Figure 2: Fraction of 1,000 random item-wise reallocations of assessments in which the maximum statistical accuracy of the scrambled panel is less than the maximum statistical accuracy in the original panel. If there were new difference between the original and scrambled panels, the values would be uniformly distributed, there would be just as much color above as below 0.5.

So, if experts quantify their uncertainty, we can combine their uncertainties and extract point predictions from the combinations, as in Figure 1. We can also quantify their performance as uncertainty assessors and use that information to form performance-weighted combinations of uncertainty. That led to a better point prediction in Figure 1. How does the method work out for all 580 assessments for which we know the true values?

Some variables may be in kilograms, others in meters per second, others in micrograms per cubic meter, etc. To compare errors in point forecasts across sets of variables with different dimensions we must make the errors dimensionless. There are many ways to do this, the most popular is to use the Mean Absolute Percentage Error (MAPE) |(prediction – realization)/realization)|. Absolute percentage errors can be added, averaged and compared across different variables. This works only if the realization (the true value) is nonzero, leaving us with 570 forecasts. Table 1 compares the Mean Absolute Percentage Errors for three expert based synthetic forecasters. Performance weighted combinations of expert distributions yields a distribution per variable whose median is taken as a point forecast (PW). Equally weighted combinations of the experts' distributions yields different medians as point forecasts (EW). Simply averaging experts' medians without regard to their uncertainties yields a third point forecast. Unwary practitioners often do the latter.

Mean Absolute Percentage Error, 49 Studies, 570 forecasts			
	PW	EW	Ave of Medians
Mean	2.2	3.8	1472.3
Stdev	11.8	45.2	33299.8

Table 1 Mean Absolute Percentage Errors for 49 studies involving 570 forecasts for performance weighting (PW) equal weighting (EW) and averaging experts' point forecasts (medians).

Simply averaging point forecasts is catastrophically bad, the average percentage error is 1,472. If the true value were 1, this would correspond to a forecast of 1,471, or -1,471. Note, this is the average percentage error over

570 forecasts.. The largest percentage error over all 570 variables in this dataset is 793,667. These errors are the result of averaging the errors over all experts in a panel.For this particular question (the estimated current global rate of magma production in the subduction zone per 1000 years)[35], one expert had a percentage error of 100 billion. The variable concerned the amount of magma released by a possible volcano. These numbers can be very small or VERY big, the realization was 6,000km^3 / 1000 yr. It is not uncommon that analysts call such large errors "outliers" and remove them to makes the error distribution "look normal". On a single data set they stick out like a sore thumb. However, when we survey a large set of errors, we realize that the distribution absolute percentage errors is very "fat tailed". The chapter on wisdom of crowds (chap 4.5) pursues this topic. We note here one characteristic of fat tailed distributions: the largest value is so much bigger than all the others that it tends to dominate the average. For this reason some analysts would prefer reporting the geometric mean which is insensitive to large numbers and sensitive to small numbers. The geometric mean[36] of PW's absolute percentage errors is 0.38 while that of averaging medians is 0.63. That might look better but ultimately we must deal with errors in the original units. Its not just a question of throwing out the largest error, fatness of tail characterizes the whole distribution, the differences between successive values at the high end keep getting larger the more observations we make. On this data set the second largest absolute percentage error is 33,914, the third is 3,091. We should have to throw out lots of forecasts to make things "look normal", because they simply are not normal.

[35] *"What is estimated to be the current global production rate of magma in subduction zone / convergent margin regions, per 1,000 years according to a global compilation of data by Crisp (1984)? (Please give your answer in km^3)".*

[36] The geometric mean of positive numbers $x_1,\ldots x_n$ is $\exp^{1/n \Sigma \ln(xi)}$.

4.4 The Iceman Cometh[37]

Date May 20, 2019 Image By Christine Zenino from Chicago, US - Greenland Ice, CC BY 2.0

The title of Eugene O'Neill's 1939 noir epic on man's need for self-deception could be the chyron for a recent article in the *Proceedings of the National Academy of Sciences* (PNAS) entitled "**Ice Sheet Contributions to Future Sea Level Rise from Structured Expert Judgement**" by J.L. Bamber, M. Oppenheimer, R.E. Kopp, W.P. Aspinall, and R.M. Cooke. Many recent publications warn of ice sheets' growing instability (here, here, and here, for example). The PNAS paper describes a structured expert judgment uncertainty quantification of ice sheets' contribution to sea level rise (SLR) out to 2300 under +2°C and +5°C stabilization scenarios. Expanding on the methodology of Bamber and Aspinall's groundbreaking 2013 study, the PNAS study again treats individual experts as testable statistical hypotheses, but this time, it targets upper-tail dependence between ice sheet processes. The result is higher median assessments and expanding uncertainties, especially in the upper tail, relative to the Fifth Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC AR5). Figure 1 shows AR5's "likely" range of 17th, 50th, and 83rd percentiles of ice sheet's contribution to sea level rise in 2100 under the Representative Concentration Pathway (RCP) 8.5 and compares those with the PNAS study, adapted to the RCP8.5 temperature trajectory (SEJ2018 ≡RCP8.5). The median 49cm (PNAS) contrasts sharply with 19cm (AR5), while the 83rd percentile 102cm (PNAS) dwarfs 35cm (AR5). An attenuated increase is found with respect to Bamber and Aspinall (2013). Due to IPCC's focus on "likely" ranges, comparisons of 5th and 95th percentiles are not possible.

[37] Based on the author's blog https://www.resources.org/common-resources/iceman-cometh/ .There is also a 15 min video on this subject by the author and a 1 hr talk for the International Geological Society by Jonathan Bamber on epistemic uncetainty in ice sheet projections.

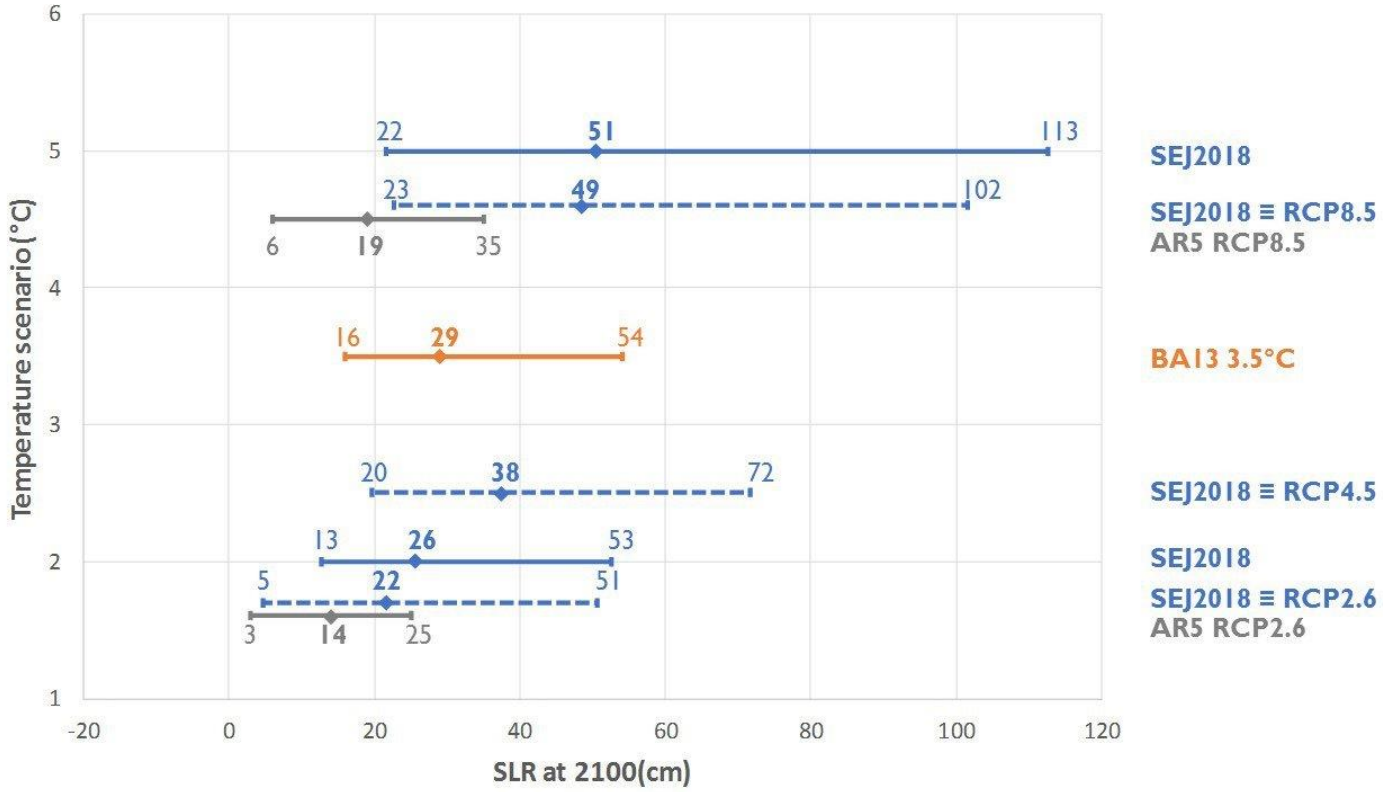

Figure 1. Median and Likely Range (17-83rd percentile as used in the AR5) Estimates of the Ice-Sheet SLR Contributions for Different Temperature Scenarios and Different Studies.AR5 RCP ice-sheet contributions are shown for RCP2.6 and 8.5 by combining contributions from the different sources (grey bars). BA13 is shown for the elicited temperature increase of 3.5° C by 2100 (orange bar). This study (SEJ2018, in blue) is shown for the Low and High temperature scenarios using solid lines. Dashed lines are interpolated from the Low and High findings using stochastic resampling of the distributions assuming a linear relationship between pairs of Low and High samples.

BOGSAT and Beyond

The contrasts in the above figure are as much about method as about numbers. The BOGSAT (bunch of guys/gals sitting around a table, 1961) approach reigns at the IPCC for uncertainty quantification. Its calibrated uncertainty language, in the form of Kent charts (1964), was eventually adopted by the US Defense Intelligence Agency in 1976, only to be abandoned a few years later for lack of validation: "there is no indication that estimates which are '70 percent probable' have been tested to determine that they were correct 70 percent of the time" (Morris and D'Amore p. 5–26).

People sitting around a table vectored to agree on an uncertainty characterization will easily focus on "likely" regions of the uncertainty space at the expense of "unlikely" regions where consensus is more difficult. Risk analysts, however, must survey the full distribution of possible outcomes from Pollyanna to Chicken Little. It is sobering to realize how large the uncertainties are. For the +5˚C stabilization scenario in 2300, the performance weighted combination of experts has a 90 percent confidence range of *minus* 9 cm SLR, due to ice sheet contributions, to +9.7 meters. Even under these conditions, it is possible—not likely but possible—that the ice sheets will *lower* the sea level by at least 9cm. It is equally likely that they will *raise* the sea level by at least 9.7 meters (these numbers are relative to sea level between 2000 and 2010 and exclude a 0.76mm/yr SLR baseline adjustment since pre-industrial times, which is present in Figure 1).

The structured expert judgment method applied in the PNAS study features traceable individual elicitations of highly vetted experts quantifying their 5th, 50th, and 95th percentiles for uncertain quantities. In addition to the variables of interest, these include variables from their field, the true values of which are known post hoc. This enables testing the hypothesis that an expert is statistically accurate—that is, statistically, 5 percent of the true values fall beneath the expert's 5th percentiles, 45 percent fall between the 5th and 50th percentiles, etc. In comparable studies, about 40 percent of experts, regarded as statistical hypotheses, would not be rejected at the 0.01 level—the same percentage as in the PNAS study. The reality is that experts are not trained in probabilistic assessment, and the majority does not perform these tasks well. Combining experts based on their performance has been shown superior to simple equal weighting both in-sample and out-of-sample.

Without a defensible quantification of uncertainty it is easy for climate derniers to lay a Confidence Trap: those favoring climate action are challenged to prove that human-caused climate change is real. This challenge is a fool's errand that must never be accepted. Indeed, if ice sheet *growth* is possible, you can't prove it isn't possible. The relevant question is not *what can we prove b*ut *what should we plan for in light of the uncertainties?* The PNAS study takes a position on this with respect to sea level rise in 2100:

"We find that, since the AR5, expert uncertainty has grown, in particular, due to uncertain ice dynamic effects . . . . For a 5°C temperature scenario more consistent with unchecked emissions growth, the [median and 95th percentile] values are 51 cm and 178 cm, respectively. Inclusion of thermal expansion and glacier contributions, results in a total SLR estimate that exceeds 2m at the 95th percentile. Our findings support the use of scenarios of twenty-first century total SLR exceeding 2m for planning purposes."

4.5 Wisdom of Crowds?

It seems to have started in 1841 with Charles Mackay's *Memoirs Of Extraordinary Popular Delusions and The Madness Of Crowds*. Francis Galton parried in 1907 with Cornwall fair goers' average (originally median) estimate of a dead bull's weight which was nearly spot on. James Surowieki's *The Wisdom of Crowds* (2004) distinguished wise crowds from irrational crowds on five criteria: diversity, independence, decentralization, aggregation and trust. Douglas Murray brought us back to *The Madness of Crowds, gender, race and identity* (2019). Much is written on the credibility /credulity of crowds. Lacking is any scientific use of **expert probabilistic** forecasting data for which realizations or true values are available. Emphasis is placed on "expert" and "probabilistic" for a number of reasons: (1) the inevitable winnowing of reliable crowds often turns on predicates associated with expertise, (2) experts' scientific training distinguishes knowledge from uncertain guesses, the provenance of forecasting, (3) probabilistic forecasting converts all quantities to a common scale, namely probability, because of which (4) we can develop performance metrics applicable to any forecast situation, and finally (5) we have extensive data from 107 structured expert judgment (*SEJ*) panels. There is even discussion whether Galton's "*vox populi*" shouldn't be called "*vox expertorum"* given the large number of expert butchers and farmers attending these events.

SEJ panels consist, on average, of *11* vetted experts giving *5, 50* and *95* percentiles for uncertain variables from their fields and also for, on average, *14* calibration variables from their fields to which true values are or become known. Performance on these calibration variables is used to construct performance weighted combinations and compared with equally weighed combinations. Expert performance is persistent, performance based combinations are superior to equal weight combinations both in– and out–of sample and have been evaluated in real applications.

Crowd–casting versus SEJ forecasting
When crowd–casting and expert forecasting mingle, Surowieki's criteria run up against expert communalities. Scientists in an SEJ forecasting panel have similar training, follow the same literature and often know each other. Physicist Max Planck (1950) famously quipped "science advances one funeral at a time". Surowieki opines: "Homogeneous groups, particularly small ones, are often victims of what the psychologist Irving Janis called "groupthink."(p.36) After a survey of expert forecasts and analyses in a wide variety of fields, Wharton professor J. Scott Armstrong wrote,"I could find no studies that showed an important advantage for expertise'(p.33). The antidote is crowd size: "...much of what we've seen so far suggests that a large group of diverse individuals will come up with better and more robust forecasts and make more intelligent decisions than even the most skilled 'decision maker'."(p.32). Au contraire, says Naomi Oreskes in *Why Trust Science* (2019): scientific consensus resulting from rigorous peer review provides a basis for trust.

SEJ data is used to examine two address two questions regarding Wisdom of Crowds (WOC): (1) Is crowd size really beneficial? and (2) Is "diversity / independence" beneficial? To address these, 40 forecasting panels with at least 10 experts and at least 10 calibration variables are selected giving 586 forecast variables with realizations, 698 experts and 10,189 expert forecasts. "Beneficial" is measured by two performance metrics.

Performance metrics
The absolute percentage error for forecast f with realization r is |(f–r)/r| and is unstable for r close to zero.Absolute percentage error is scale invariant, so scores for different forecasts and different realizations can be averaged, yielding the Mean Absolute Percentage Error (MAPE). We can average over all calibration

variables for each expert to form an expert specific MAPE. We can average the expert specific MAPEs for all experts in a panel to arrive at a panel specific MAPE which is the expected MAPE of a randomly chosen expert. Invoking the Wisdom Of Crowds (WOC) we can first average experts' median forecasts and then compute the WOC MAPE, per variable and per panel. A well-known mathematical result (Jensen's inequality) says that WOC MAPE is always less or equal to expert panel MAPE (i.e. better), though the mean difference per panel is a mere 0.009 in this case. Yes, WOC-ing helps but not very much.

Statistical Accuracy (SA) is based on the relative frequency with which the realizations of independent calibration variables fall inside the forecaster's four inter–quantile intervals. SA is the probability that these relative frequencies should differ from the theoretical inter quantile probabilities (5%, 45%, 45%, 5%) by at least the observed amount. Low values near zero mean that it is very unlikely that the forecaster's probabilities are statistically accurate, high values, near 1, indicate good agreement between observed and expected relative frequencies.

When we assemble a panel of experts we are in fact drawing a small sample from a large set of potential experts. We now have the ability to study this large set.

Tail Size[38]

Participants in WOC discussions need to appreciate how much the discussion has been constrained by statistical assumptions, and how fragile these assumptions really are. If we sample a set of numbers from some distribution, we can always compute the average of these numbers as well as the variance, standard deviation, correlations with other sets of numbers etc. But if we sample more numbers or sample a like sized second batch, do these averages, variances and correlations tend to agree? The law of large numbers says that averages, variances, and correlations stabilize as we draw ever larger samples; however this law applies only if the distribution from which the numbers are drawn is "thin tailed". If the distribution is "fat tailed" then none of this holds.

Pictures give a better idea than formal mathematical definitions. The left panel of Figure 1 gives running averages (average the first two, then the first three, etc) of 1000 independent samples from a uniform distribution on the [0, 1] interval. The horizontal axis gives the size over which the average is taken. On the vertical we plot the running average up to the corresponding size. At the horizontal value 1000 we average all 1000 samples. In the right panel we do the same, with the same numbers, except that these numbers are now inverted; 0.1 becomes 10, etc. The distribution of the inverse of the uniform variables is a very fat tailed distribution. With thin tailed distributions running averages converge, with fat tailed distributions they do not. Very large values keep popping up at a rate which prevents convergence. Fat tailed distributions are common, but not common knowledge.

Statisticians don't like fat tails, as they prevent application of the familiar statistical methods. It's easy to delete a single large value as an "outlier" so that the rest "look normal". However, when one looks at larger samples from a fat tailed distribution, one realizes that the large values are characteristic of the whole distribution. If we order the sample from smallest to largest values, we see that the distance between adjacent samples just gets larger as the sample values get larger. As we gather more samples, the average of the whole sample tends to resemble the largest sample in the set. In fact, this is a defining feature of the "subexponential" class of fat tailed distributions.

[38] This is a non-technical introduction to fat tailed distributions. Many text books give a full mathematical treatment, Cooke et al (2014) is directed to numerate non-specialists.

Figure 1: Thin tailed (left) and fat tailed (right) running averages. Horizonal axes denote the number of samples over which we average.

If we draw repeated samples of size 1000, the thin tailed running averages will differ a bit at the beginning but quickly settle into the pattern. Figure 2 shows what happens with three samples of 1000 from the from the distributions in Figure 1. Notice the changing scale on the vertical axis for the fat tailed distributions; these samples do not settle into a pattern. The are dissipative. The different pictures depend on where in the list of 1000 variables the largest value ocurrs, near the beginning (bottom right) or near the end (bottom middle) or in between (bottom left).

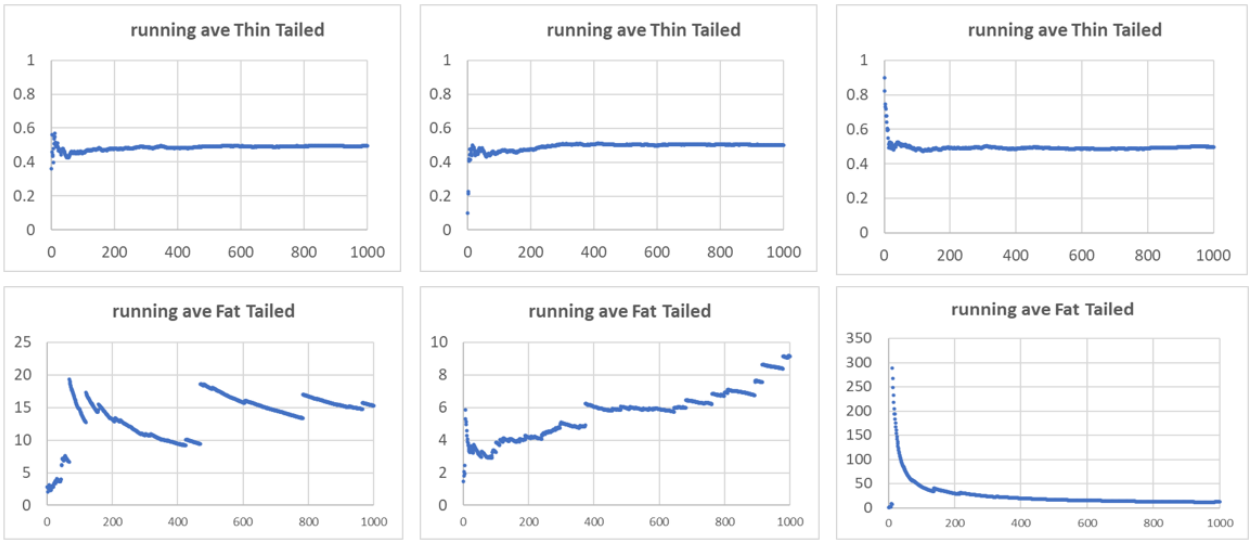

Figure 2: Three Repeated samples with running averages

Once we look, we can see fat tailed distributions everywhere - damages from natural disasters, crop insurance claims, citation scores, flood damages, income distributions, hospital discharge rates etc. (Cooke et al 2014). What about experts?

Crowd size

Do averages of ever more forecast errors trend down? There is an antecedent question: Do such averages converge at all? Figure 3 shows running averages of US damages in excess of 10$M due to natural disasters (left, Cooke et al 2014) and absolute percentage error in 10,198 expert forecasts (right). To be sure, these experts assess different quantities, but their absolute percentage errors can be plotted on an absolute scale reflecting the factor by which the forecast differs from the realization in absolute value. Such graphs depend on

the ordering but random re–orderings will exhibit the same key feature: ever larger values keep popping up that prevent convergence.

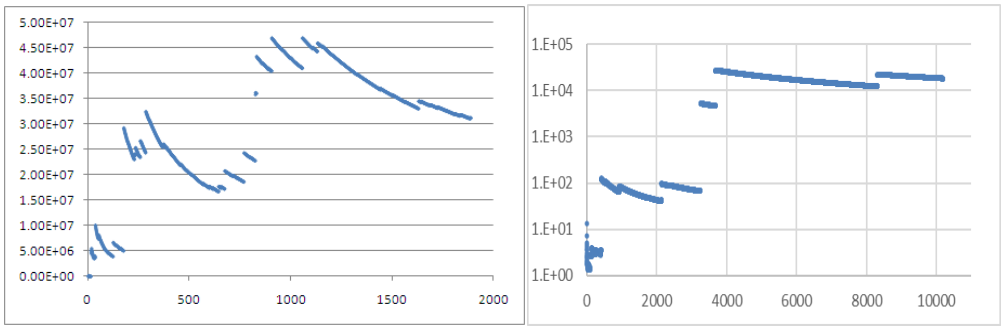

Figure 3: Running averages for US damages in excess of 10$M (right, Cooke et al 2014) and running averages for 10,189 expert absolute percentage errors.

Experts' absolute percentage errors in aggregate are very fat tailed. Does this also apply to WOC forecasts where expert forecasts are averaged per panel? Figure 4 (left) shows 586 realizations in ascending order plotted with their WOC forecasts. Note the very small realization with forecast differing by 8 orders of magnitude.To avoid instabilities due to small realizations, we subset the 535 forecasts for which the realizations are greater or equal to 0.1. The running averages are shown in Figure 2 (right); the averages of ever larger sets of WOC absolute percentage forecast errors just keeps growing. This explains the very large errors in Table 1 of chap. 4.3.

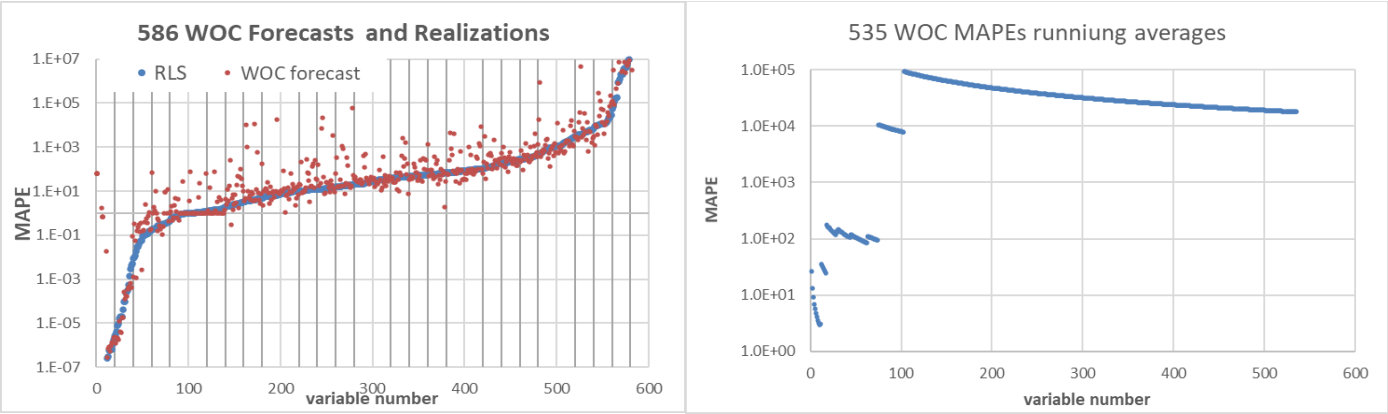

Figure 4: 586 WOC forecasts against realizations (left) and running averages of 535 WOC MAPEs (right)

The panel sizes in our dataset do not support tail analysis per panel, but the effect of number of forecasters can be seen in other ways. Figure 5 plots all 586 WOC MAPEs against the panel size over which the median forecasts are averaged. The rank correlation in Figure 3 is weakly positive. WOC panel MAPES are not decreasing in panel size.

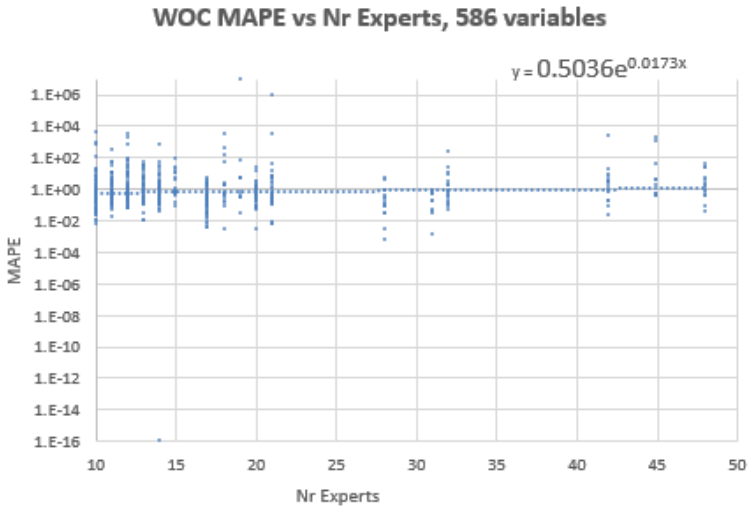

Figure 5: WOC MAPE against number of empaneled experts for 586 variables.

The question of crowd size is related to the question of diversity: if all experts say the same thing then crowd size doesn't matter. There are various ways to measure diversity in a panel of experts, a technical discussion is available at http://rogermcooke.net/. Suffice to say that we can measure the disagreement in an expert panel in two ways, disagreement with respect to point forecasts and disagreement with respect to 90% uncertainty bands. Experts tend to cluster moderately with regard to the point forecasts, but much less so regarding the 90% confidence bands. The more they agree on the point forecasts, the less diverse they are and the *better* is the performance of the WOC average forecast – exactly the opposite of what the WOC'ers suppose. The reason is simple: Since the point forecast absolute percentage errors are very fat tailed, the average percentage error gets larger as we enlist more experts. A decrease of diversity reduces the effective crowd size and acts as a break on the increase of percentage errors. At the same time, more DISagreement regarding point forecasts tends to improve the mean statistical accuracy of the panel, reinforcing the diversity theme. Note that measuring statistical accuracy requires the experts' uncertainty bands, it cannot be done using only the point forecasts. A simplistic rendering is: diversity is good for statistical accuracy, and bad for point forecast percentage errors. In short, its complicated. Absent good expert data there is very little chance of getting it right.

Two over-arching conclusions emerge: (1) Averaging experts' point forecasts is catastrophically bad. Forming point forecasts from medians of equally weighted combination of experts' distributions is much better, and using performance weighted combinations is better still. (2) Performance weighting tends to reduce the effective panel size by down selecting high performers and this suppresses the very large percentage errors.

5. The Hard Part: Twining

The threads must come together. Lets take a step back. To the ancient Greeks a number was a ratio of comensurable quantities, one can divide meters by meters, seconds by seconds, to get numbers. How could one divide meters by seconds? It took mankind about 1500 years to learn how divide meters by seconds and arrive at the notion of velocity (Dijksterhuis 1961). Russell's paradox[39] showed that the naïve notions of set theory in common use were inconsistent. Researchers sought a set of axioms which was adequate for everyday mathematics and provably consisternt. The adequacy requirement was met[40] (Fraenkel et al 1973) but Gődel (1931) showed that any such system could not be proved consistent. Exotic variants like non-standard analysis and intuitionism appeared, flourished for a time, then receded. In logic, the first order predicate logic became the logic of set theory, adequate for every day mathematics. Exotics like multi-valued logic, higher order logic, fuzzy logic and quantum logic emerged, proliferated and waned. The pattern repeats itself for uncertainty and for the axioms of rational decision theory. Throughout all this, the debate is not over which axioms are TRUE. Exotics are not abandoned because they are false; indeed, they may come back. However, core theories eventually emerge which are adequate, easy to understand and, especially, easy to explain. The exotics become variations on the core theme. New problems emerge and we move on: Passez outre….next myth.

The philosopher G.W. F. Hegel never wrote an ethics, instead we have Hegel's Philosophy of Right, or Philosophy of the State[41]. The State's authority is grounded not on Natural Right, not on a Social Contract, not on the Right of the Strongest, not on Tradition, but on the Universal Recognition of the Right to Property: Every person recognizes the right of every other person to put his/her mark on a piece of unclaimed nature and declare it his/her property, to dispose of as (s)he pleases. Hegel's State emerges, - finally, at the end of history - as expression of, and the guarantor of the universal right to property. This right immediately entails the universal right to life and prevents a person from being some one else's property.

The Universal Right to Property is very different from the Absolute Right to Property, according to which I can take possession of anything I like, including you, and do with it as I please; if you don't like it, you can try to take it back. This distinction is alas lost on our libertarian friends. According to MacLean (2023) the orgin of libertarianism and its war on taxation dates back to South Carolina senator and slavery advocate John C. Calhoun (1782-1850). Economics professors at Charles Koch's operation at George Mason University saw in him "a precursor of modern public choice theory", who were convinced that democracy could not "preserve liberty" or protect the "tax producers from the "tax consumers". MacLean, Nancy. Democracy in Chains: The Deep History of the Radical Right's Stealth Plan for America (p. 1). Penguin Publishing Group. Kindle Edition. An early advocate was University of Chicago professor Frank Knight of Knightian uncertainty (see Chapt 3.10) who inspired libertarians James McGill Buchanan and Milton Friedman, among many others. Their "market order" cures all ills, including racism and, (by presumption, slavery, child labor, gender discrimination, anti semiticism and the rest?). McLean relates::

> "*On a visit to Harvard, Friedman devoted most of his speech to criticizing the Civil Rights Act, complaining that it used "coercive" means to make all "conform to the values of the majority," in violation of the liberty of the white minority that opposed reform. Friedman urged reliance instead on "free market principles": prejudice would cause lower wages for black workers, which in turn would*

[39] Let A be the set of all sets which are not members of themselves. If A is a member of itself, then it is not, and if it is not, then it is.

[40] The Zermelo-Fraenkel axioms with the axiom of choice describe the actions one can perform on sets to produce new sets

[41] When referring to the State as the residence of social authority, the upper case is used, lower case is used for other states of affair.

> *reduce production costs for those who employed them, so more employers would hire African Americans, he said—and, presto, "virtue triumph[s]."*[10] MacLean, Nancy. Democracy in Chains: The Deep History of the Radical Right's Stealth Plan for America (p. 90). Penguin Publishing Group. Kindle Edition.

Do we need property and a State at all? The inevitability of the State was not apparent to our Wendat philosopher Kandiaronk as he told his French colonialist friends : "*For my own part, I find it hard to see how you could be much more miserable than you already are. What kind of human, what species of creature, must Europeans be, that they have to be forced to do good, and only refrain from evil because of fear of punishment?… You have observed that we lack judges. What is the reason for that? Well, we never bring lawsuits against one another. And why do we never bring lawsuits? Well, because we made a decision neither to accept or make use of money. And why do we refuse to allow money into our communities? The reason is this: we are determined not to have laws – because, since the world was a world, our ancestors have been able to live contentedly without them.*" Graeber, David. The Dawn of Everything: A New History of Humanity (p. 54). Farrar, Straus and Giroux. Kindle Edition.

The recognition of the Universal Right to Property is just the beginning. On page 47 Hegel writes (good luck with this):

"Man pursuant to his immediate existence within himself, is something natural, external to his concept. It is only through the development of his own body and mind, essentially through his self consciousness's apprehension of itself as free, that he takes possession of himself and becomes his own property and no one else's. This taking possession of oneself, looked at from the opposite point of view, is the translation into actuality of what one is according to one's concept, i.e. a potentiality, capacity, potency. In that translation one's self-consciousness for the first time becomes established as one's own, as one's object also and distinct from self-consciousness pure and simple and thereby capable of taking the form of a 'thing'" (p.47-8).

Man takes possession of him/herself in and by owning him/herself, to do with as (s)he pleases. Self possession is, if you will, the Will applied to itself.

" …that man's absolute unfitness for slavery should no longer be apprehended as a mere 'ought to be', is something which does not come home to our minds until we recognize that the Idea of freedom is genuinely actual only as the State"(p. 48)

Self possession pivots the narrative from "Abstract Right" to "Morality" and on to "Ethical Life ("Sittlichkeit")" which is "*the good endowed with self-consciousness with knowing and willing and actualized by self-conscious action..:"* (p. 105). Freedom is only actualized as the State.

How does self-possession emerge as ethical life? Myths and allegories have a venerable history in philosophy, Zeno's Tortoise and Hare, Plato's Allegory of the Cave, Pascal's Wager, Rawls' Original Position. The answer to this question and conclusion of this essay is a Hegelian myth.

Agent Based Model of Inhering[42] *Value*

[42] "You're probably familiar with *inherent*, the adjective meaning "part of the constitution or natural character of something," but were you aware of its less common relative *inhere*? This verb looks like it could be a back-formation of *inherent* (a back-formation is a word created by removing a prefix or suffix from an existing word), but usage evidence of the two words makes it difficult to tell for sure. " https://www.merriam-webster.com/dictionary/inhere

Things acquire instrumental value if they assist in securing other things of value. Inherent value is pursued not in order to obtain something else; its possession is immediately valuable in and of itself. The agent with inhering value is an android with the following features

1. A powerful AI engine capable of forming beliefs, choosing and executing actions to secure its necessary resources (energy, maintenance etc). In so doing, the AI engine learns about the surrounding world with which it interacts.
2. An affective scale representing the desirability of its current state. Position on this affective scale is influenced by the agent's actions and by exterior forces which the AI engine does not directly control. It is a simple scale taking values from zero to one. One, the highest affective state, is termed the *summum bonum*, and the lowest state, zero, is termed *pessimus malum*. Think of it as digital dopamine.
3. A purpose subroutine: it is programmed to formulate and perform actions to raise its affective state. Its affective state changes as a result of its actions and other exterior influences outside its direct control.

The agent is a sheer libertarian attending only to its own affective state; yet it contains the potentiality for intrinsic value in the following sense: When the agent 'can do with itself what it pleases', when it 'possesses itself' , when it can choose its own affective state, it puts itself, by definition, in the state of *summum bonum*. The set {1,4,9} and the set {perfect squares from 1 to 9} are co-terminal. They denote the same thing albeit in different ways. The notions of full self possession and *summum bonum* are co-terminal, they denote the same thing in different ways. Similarly, the state of *pessimus malum* is co-terminal with the state of complete self-dispossession, or total loss of agency. In *pessimus malum* no action could make the agent more miserable, the agent would do anything to be in another affective state, by definition. By extension, intermediate affective states are co-terminal with degrees of self possession. These are intrinsic values, (un)desirable in themselves not by dint of relation to other values. Understood in this way, the immediacy of intrinsic value is tautological, true by definition. The agent doesn't interact with the exterior to learn its own affective state, it is not uncertain which affective state it is in.

Notice also this: the agent does not learn its affective from its AI engine through external interactions. Similarly, in a state of euphoria we do not question whether we are euphoric, in despondency we are unquestionably despondent. These are not states of the world about which we can be uncertain, but states of the acting subject. In Savage's rational decision theory actions are utility-valued functions on the set of possible worlds; we are uncertain about states of the world but not about our utilities.

In the myth so far there is intrinsic value but no State. We now endow the agent with another property:

4. The agent's affective state resonates to varying degrees with changes in the affective states of other agents, if such there be. If the agent's actions get associated with an elevation in the affective state of another agent, that feeds back to some degree to agent's own state inducing a resonant elevation, idem degradation. This feed back is empathy – the feeling of another's pain or joy. The AI engine is not learning the other's affective state, but feeling it, as it were. It does learn that raising (lowering) another's affective state raises (lowers) its own.

Empathy is a way of transferring the affective state of one agent to another. As a set of agents develop common beliefs, formulate common actions and become empathetic, the set of agents itself begins to satisfy properties (1,2,3) and itself becomes a value inhering agent: it becomes a State. Sets of States can do the same…. The rest is – unfinished – history.

References

Allais, M. (1953) Le comportement de l'homme rationnel devant le risque. Econometrica 21, 503–546.

Archer, D. (2016) The Long Thaw: How Humans Are Changing the Next 100,000 Years of Earth's Climate, Princeton Uiniversity Press.

Aspinall, W.P. (2010) "A route to more tractable expert advice" Nature, vol. 463, 21 January, 2010.

Aspinall, W.P., Cooke, R.M. (2013) Quantifying scientific uncertainty from expert judgement elicitation. In "Risk and Uncertainty assessment in Natural Hazards." Hill, L., Rougier, J.C., Sparks R.S.J. (eds). Cambridge University Press, Chapter 4, 64-99. ISBN: 9781107006195

Aumann HH et al (2003) AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems. IEEE Trans Geosci Remote Sens 41(2):253–264. https://doi.org/10.1109 /TGRS.2002.808356

Bamber, J.L., and Aspinall, W.P., (2013) An expert judgement assessment of future sea 1evel rise from the ice sheets, Nature Climate Change,PUBLISHED ONLINE: Janurary 6, 2013 | DOI: 10.1038/NCLIMATE1778.

Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F.R.S., communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S. Philosophical Transactions of the Royal Society of London 53, 370–418. http://rstl.royalsocietypublishing.org/content/53/370.full.pdf+html (accessed April 16, 2014).

Brown PT, Li W, Cordero EC, Mauget SA (2015) Comparing the model-simulated global warming signal to observations using empirical estimates of unforced noise. Nat Sci Rep. https://doi.org/10.1038/srep09957

Cooke RM (1983) A result in Renyi's conditional probability theory with application to subjective probability. J Philos Log 12(1):19–32 http://www.jstor.org/stable/30227015

Cooke RM, Golub A, Wielicki BA, Young DF, Mlynczak MG, Baize R (2016) Real option value for new measurements of cloud radiative forcing, Resources for the Future, RFFDP 19–16, March 22, 2016

Cooke RM, Golub A, Wielicki BA, Young DF, Mlynczak MG, Baize R (2015) Integrated assessment modeling of value of information in earth observing systems. Clim Pol ISSN: 1469–3062 (print) 1752–7457 (online) journal homepage: http://www.tandfonline.com/loi/tcpo20

Cooke R.M., Wielicki BA, Young DF, Mlynczak MG (2013) Value of information for climate observing systems. Environ, Syst Decis. https://doi.org/10.1007/s10669-013-9451-8

Cooke, R.M. (1991) Experts in Uncertainty: Opinion and Subjective Probability in Science. New York, NY: Oxford University Press.

Cooke, R.M. (2003) Elicitation of expert opinions for uncertainty and risks. Fuzzy Sets and Systems 133, 267–268.

Cooke, R.M. (2008) Response to comments. Reliability Engineering & System Safety 93 (5), Special issue on expert judgment, 775–777.

Cooke, R.M. (2013) Uncertainty analysis comes to integrated assessment models for climate change … and conversely. Climatic Change 117 (3), 467–479. http://dx.doi.org/10.1007/s10584-012-0634-y (accessed April 16, 2014).

Cooke, R.M., ElSaadany, S., and Huang, X. (2008) On the performance of social network and likelihood based expert weighting schemes. Reliability Engineering & System Safety, 93 (5), Special issue on expert judgment, 745–756.

Cooke, R.M., Goossens, L.H.J. (2008) TU Delft Expert Judgment Data Base. Reliability Engineering & System Safety, 93 (5), Special issue on expert judgment, 657–674.

Cooke, R.M., Nieboer, D. Misiewicz, J. (2014) Fat Tailed distributions, data, diagnostics and dependence, Wiley, London. https://www.wiley.com/en–be/Fat+Tailed+Distributions:+Data,+Diagnostics+and+Dependence,+Volume+1–p–9781848217928

Cooke, Roger M. (2014) Validating Expert Judgments with the Classical Model" in Experts and Consensus in Social Science - Critical Perspectives from Economics, Sociology, Politics, and Philosophy. Editors: Carlo Martini and Marcel Boumans, Series title: Ethical Economy - Studies in Economic Ethics and Philosophy, Springer

Cooke, R.M. & Wielicki, B. (2018) Probabilistic reasoning about measurements of equilibrium climate sensitivity: combining disparate lines of evidence, Climatic Change, 151(3), 541-554 https://doi.org/10.1007/s10584-018-2315-y

Cooke, Roger M., Marti, Deniz and Mazzuchi, Thomas A., (2021) Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say? International Journal of Forecasting, published online July 25, 2020, https://doi.org/10.1016/j.ijforecast.2020.06.007 .

Cooke R.M. (2021) Building on Foundations: an interview with Roger Cooke, in Expert Judgement in Risk and Decision Analysis eds Nane, Hanea, French and Bedford, Springer Nature Switzerland AG, Cham, Switzerland.

DeGroot, M. (1974). Reaching consensus, J. Amer. Statis. Assoc. vol.69, pp118 - 121

Dessler AE (2010) A determination of the cloud feedback from climate variations over the past decade. Science 330:1523–1527. https://doi.org/10.1126/science.1192546

Dessler AE, Forster PM (2018) An estimate of equilibrium climate sensitivity from interannual variability. J Geophys Res Atmos 123. https://doi.org/10.1029/2018JD028481

Dessler AE, Mauritsen T, Stevens B (2018) The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. Atmos Chem Phys 18:5147–5155. https://doi. org/10.5194/acp-18-5147-2018

Dijksterhuis, E. J. (1961). The mechanization of the world picture. London: The Oxford University Press.

Eggstaff, J.W., Mazzuchi, T.A., and Sarkani, S. (2014) The effect of the number of seed variables on the performance of Cooke's classical model. Reliability Engineering & System Safety 121, 72–82.

Ellsberg, D. (1961) Risk, ambiguity, and the Savage axioms. The Quarterly Journal of Economics 75, 643–669.

Embrechts, P., Kuppelberg, C. and T. Mikosch. Modelling Extremeal Events for Insurance and Finance. Springer, 1997.

Evans, D.J., and Sezer, H., (2005), Social discount rates for member countries of the European Union. Journal of Economic Studies 32 (1), 47–59.

Feigenbaum, E.A. (1977) The art of artificial intelligence: I. Themes and case studies of knowledge engineering. In Proceedings of the International Joint Conference on Artificial Intelligence and the National Computer Conference. http://ijcai.org/Past%20Proceedings/IJCAI-77-VOL2/PDF/092.pdf (accessed April 16, 2014). Also published as Stanford Heuristic Programming Project Memo (112) HPP-77-25, and Computer Science Department Memo STAN-CS-77-621.

Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., and Ginzburg, L. (2007) Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. SAND2007-0939. Albuquerque, NM: Sandia National Laboratories.

Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, 1535 pp. https://doi.org/10.1017/CBO9781107415324

Fraenkel, A.A., Bar-Hillel, Y. and Levy, A. (1973) Foundations of Set Theory, North-Holland, Amsterdam.

Fredrick, S., Loewenstein, G., and O'Donoghue, T. (2002) Time discounting and time preference: A critical review. Journal of Economic Literature 40 (2), 351–401.

Frege, Gottlob (1953) The Foundatijons of Arithmetic, A logico-mathematical enquiry into the concept of number, Harper Bros, New York,, first published 1884.

Fu, G., Hall, J., and Lawry, J. (2005) Beyond probability: New methods for representing uncertainty in projections of future climate. Working paper 75. Norwich, UK: Tyndall Centre for Climate Change Research.

Ghosh, S., and Mujumdar, P.P. (2009) Climate change impact assessment: Uncertainty modeling with imprecise probability. Journal of Geophysical Research 114, D18113, doi:10.1029/2008JD011648.
Gődel, K. (1932) On formally undedidable propositions of Principia mathematica and related systems I,(1931) in Frege and Gődel, Two Fundamenta Texts in Mathematical Logic, Jean van Heijenoort (ed), Harvard University Press, Cambridge, 1967.
Graeber, David and Wengrow, David (2021) The Dawn of Everything: A New History of Humanity, Farrar, Straus and Giroux.
Hall, J.W., Fu, G., and Lawry, J. (2007) Imprecise probabilities of climate change: Aggregation of fuzzy scenarios and model uncertainties. Climatic Change 81 (3–4), 265–281.
Hanea AM, Morales Napoles O, Ababei D (2015) Non-parametric Bayesian networks: improving theory and reviewing applications. Reliab Eng Syst Saf. https://doi.org/10.1016/j.ress.2015.07.027 0951–8320/& 2015
Hanea AM, Nane GF, Wielicki BA, Cooke RM (2018) Bayesian networks for identifying incorrect probabilistic intuitions in a climate trend uncertainty quantification context. Risk Research pp 1–16. https://doi. org/10.1080/13669877.2018.1437059
Hanea, A.M., (2008) Algorithms for non-parametric Bayesian Nets Phd Thesis Department Of Mathematics, Delft University Of Technology
Hegel, G.W. F. (1821) Naturrecht und Staatswissenschaft im Grundrisse , Grundlinien der Philosophie des Rechts, translated by T.M. Knox as Hegels's Philosophy of Right. Clarendon Press, 1967
Henry, C. 2006. Decision-Making Under Scientific, Political and Economic Uncertainty. Chaire Développement Durable, Cahier DDX–06–12. Paris, France: Laboratoire d'Econométrie de l'École Polytechnique.
Hilton F et al (2012) Hyperspectral Earth observations from IASI: five years of accomplishments. Bull Am Meteorol Soc 93:347–370. https://doi.org/10.1175/BAMS-D-11-00027
IPCC (2012) Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties, Jasper Ridge, CA, 6–7 July 2010. Geneva, Switzerland: Intergovernmental Panel on Climate Change.
IPCC (2013) Climate Change 2013: The Physical Science Basis. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Contribution of Working Group I to the
IWGSCC (Interagency Working Group on Social Cost of Carbon) (2009) Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866, Appendix 15a. US Government, Washington, DC, p 53
IWGSCC (InteragencyWorking Group on Social Cost of Carbon) (2013) Technical support document: technical update of the social cost of carbon for regulatory impact analysis under executive order 12866. US Government, Washington, DC May 2013, revised Nov 2013
IWGSCC (InteragencyWorking Group on Social Cost of Carbon) (2016) Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis
Under Executive Order 1286
Kahneman D (2011) Thinking, fast and slow. Farrar, Straus and Giroux, New York National Academies of Sciences, Engineering, and Medicine (2017) Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide. The National Academies Press, Washington, DC. https://doi.org/10.17226/24651
Kahneman D., and Tversky, A. (1979) "Prospect theory: an analysis of decision under risk", Econometrica, 47, 263-291.
Kandlikar, M., Risbey, J., and Dessai, S. (2005) Representing and communicating deep uncertainty in climate-change assessments. C.R. Geoscience 337, 443–455.
Kendall, M. G., Kendall Sheila F. H. and Babington Smith B, (1939) The Distribution of Spearman's Coefficient of Rank Correlation in a Universe in which all Rankings Occur an Equal Number of Times: Biometrika , Jan., 1939, Vol. 30, No. 3/4 (Jan., 1939), pp. 251–273 Published by: Oxford University Press on behalf of Biometrika Trust, Stable URL: https://www.jstor.org/stable/2332649
Keynes, John Maynard (1921), Treatise on Probability, London: Macmillan & Co
Knight, F.H. (1921). Risk, Uncertainty, and Profit. New York, NY: Harper & Row. http://www.econlib.org/library/Knight/knRUPCover.html (accessed April 16, 2014).
Kriegler, E., Hall, J.W., Held, H., Dawson, R., and Schellnhuber, H.J. (2009) Imprecise probability assessment of tipping points in the climate system. PNAS 106 (13), 5041–5046.
Laplace, P.S. (1814; English edition 1951) A Philosophical Essay on Probabilities. New York, NY: Dover Publications.
Lehrer, K., and Wagner, C. (1981). Rational Consensus in Science and Society, D. Reidel, Dordrecht.
Leroy SS, Anderson JG, Ohring G (2008) Climate signal detection times and constraints on climate benchmark accuracy requirements. J Clim 21:184–846
Lin, S.-W., and Bier, V.M. (2008) A study of expert overconfidence. Reliability Engineering & System Safety 93 (5), 775–777.
Lin, S.-W., and Cheng, C.-H. (2008) Can Cooke's model sift out better experts and produce well-calibrated aggregated probabilities? In Proceedings of the 2008 IEEE IEEM. Department of Business Administration, Yuan Ze University, Chung-Li, Taiwan.
Lin, S.-W., Cheng, C.-H. (2009) The reliability of aggregated probability judgments obtained through Cooke's classical model. Journal of Modelling in Management 4 (2), 149–161.
Lin, Shi-Woei, Huang, Ssu-Wei (2012) "Effects of Overconfidence and Dependence on Aggregated Probability Judgments" Journal of Modelling in Management Volume 7 issue 1
Mach, K.J., Matschoss, P.R., Plattner, G. K., Yohe, G.W., and Zwiers, F.W. (2010) Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change
Mackay, Charles (1841) Memoirs Of Extraordinary Popular Delusions and The Madness Of Crowds. Vol. I (1 ed.). London: Richard Bentley.
Mastrandrea, M., Mach, K., Plattner, G.-K., Edenhofer, O., Stocker, T., Field, C., Ebi, K., Matschoss, P. (2011) The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. Climatic Change 108, 675–691. http://www.climate.unibe.ch/~stocker/papers/mastrandrea11cc.pdf (accessed April 16, 2014).
McCrimmon, K.R. (1968). Descriptive and normative implications of the decision-theory postulates. In Risk and Uncertainty. London, UK: MacMillan, 3–24.
MacLean, Nancy. (2023) Democracy in Chains: The Deep History of the Radical Right's Stealth Plan for America (p. vii). Penguin Publishing Group. Kindle Edition.
Morgan, G.M., Dowlatabadi, H., Henrion, M., Keith, D., Lempert, R., McBride, S., Small, M., and Wilbanks, T. (2009) Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decisions. Synthesis and Assessment Product 5.2. Washington, DC: US Climate Change Science Program.
Morgan, M., and Henrion, M. (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York, NY: Cambridge University Press.
Morris, P. (1977) Combining expert judgments: A Bayesian approach. Management Science 23 (7), 679–693.
Moss, R.H., and Schneider, S.H. (2000) Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC. Pachauri, R., Taniguchi T., and Tanaka, K. (eds.). Geneva, Switzerland: World Meteorological Organization, 33–51. http://www.ipcc.ch/pdf/supporting-material/guidance-papers-3rd-assessment.pdf (accessed April 16, 2014).
Murray, Douglas (2019) The Madness of Crowds, gender, race and identity (2019), https://www.amazon.com/s?k=The+Madness+of+Crowds%2C+gender%2C+race+and+identity&i=stripbooks&crid=MY0QAN00KUHA&sprefix=the+madness+of+crowds%2C+gender%2C+race+and+identity+%2Cstripbooks%2C192&ref=nb_sb_noss
Newell, R.G., and Pizer, W.A. (2003) Discounting the distant future: How much do uncertain rates increase valuations? Journal of Environmental Economics and Management 46, 53–71.

Nordhaus W, Sztorc P (2013) DICE 2013R: Introduction and user's manual (2nd ed.). http://www.econ.yale edu/~nordhaus/homepage/documents/DICE_Manual_103113r2.pdf
Nordhaus, W. D. (2008). A Question of Balance: Weighing the Options on Global Warming Policies. New Haven, CT: Yale University Press.
Nordhaus, W., and Popp, D. (1996) What is the value of scientific knowledge: An application to global warming using the PRICE model. Discussion paper no.1117. New Haven, CT: Cowles Foundation for Research in Economics, Yale University.
Neslo R.E.J, and Cooke, Roger M.,(2011) "Modeling and Validating Stakeholder Preferences with Probabilistic Inversion" Applied Stochastic Models in Business and Industry. Applied Stochastic Models in Business and Industry Special Issue: Games and Decisions in Risk and Reliability Analysis,Volume 27, Issue 2, pages 115–130, March/April 2011. https://doi.org/10.1002/asmb.888
NRC (2010) Advancing the Science of Climate Change. Washington, DC: National Research Council. http://www.nap.edu/catalog.php?record_id=12782 (accessed April 15, 2014).
Oppenheimer M, Little CM, Cooke RM (2016) Expert judgment and uncertainty quantification for climate change, appearing in Nature Climate Change. 6:445–451. https://doi.org/10.1038/NCLIMATE2959
Oppenheimer M, O'Neill B (2008) C. and Webster. M. (2008) Negative learning. Clim Chang 89:155–172. https://doi.org/10.1007/s10584-008-9405-1
Oreskes N, Conway EM (2010) Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming, 1st U.S. edn. Bloomsbury Press, New York
Oreskes, Naomi (2019) Why Trust Science, Princeton University Press https://press.princeton.edu/books/hardcover/9780691179001/why–trust–science
Pizer, W.A. (1999) The optimal choice of climate change policy in the presence of uncertainty. Resource and Energy Economics 21, 255–287.
Planck, Max K. (1950). Scientific Autobiography and Other Papers. New York: PHiLosophical library.
Quiggin,J. (1993) Generalized Expected Utility Theory – the Rank-Dependent Model Kluwer, Dordrecht.
Ramsey, F.P. (1931; originally published 1926) Truth and probability. In The Foundations of Mathematics and Other Logical Essays. Braithwaite, R.B. (ed.). London, UK: Kegan, Paul, Trench, Trubner & Co.; New York, NY: Harcourt, Brace and Company, 156–198.
Rennert, Kevin, Frank Errickson, Brian C. Prest, Lisa Rennels, Richard G. Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke,Bryan Parthum, David Smith, Kevin Cromar, Delavane Diaz, Frances Moore, Ulrich K. Müller, Richard Plevin, Adrian E. Raftery, Hana Ševčíková, Hannah Sheets, James H. Stock, Tammy Tan, Mark Watson, Tony Wong, David Anthoff (2022) Comprehensive Evidence Implies a Higher Social Cost of CO2. Nature. https://doi.org/10.1038/s41586-022-05224-9
Rennert, Kevin, Prest, Brian, Pizer, William, Newell, Richard, Anthoff, David, Kingdon, Cora, Rennels, Lisa, Cooke, Roger, Raftery, Adrian E. Ševčíková, Hana and Errickson, Frank (2021) The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates with online appendix, Brookings Papers on Economic Activity Working Paper (21-28) Sept 9, 2021. https://www.brookings.edu/bpea-articles/the-social-cost-of-carbon/
Renyi A (1970) Theory of probability, North-Holland Soden BJ, Held IM, Colman R, Shell KM, Kiehl JT, Shields CA (2008) Quantifying climate feedbacks using radiative kernels. J Clim 21:3504–3520
Ryan, J.C.H. Mazzuchi,T.A. Ryan,D.J., Lopez de la Cruz, J. and Cooke, R. M. (2012) "Quantifying information security risks using expert judgment elicitation", Computers & Operations Research Journal 39, 774–784. , doi:10.1016/j.cor.2010.11.013
Savage, L. J., (1972) The Foundations of Statistics, Dover, New York, first published in 1954, John Wiley and Sons.
Schmeidler, D. (1989) "Subjective probability and expected utility without additivity" Econometrica 57, 571-587.
Stern, N. (2008). The economics of climate change. American Economic Review 98 (2), 1–37.
Strow LL et al (2013) Spectral calibration and validation of the cross-track infrared sounder on the Suomi NPP satellite. J Geophys Res Atmos 118:12 486–12 496. https://doi.org/10.1002/2013JD020480
Surowieki's, James, (2019) The Wisdom of Crowds (2005), Anchor; Reprint edition (August 16, 2005) https://www.amazon.com/Wisdom–Crowds–James–Surowiecki/dp/0385721706
Tversky, A., and Kahneman, D. (1982) Judgments of and by representativeness. In Judgment under Uncertainty: Heuristics and Biases. Kahneman, D., Slovic, P., and Tversky, A. (eds.). Cambridge, UK: Cambridge University Press, 84–98.
US Nuclear Regulatory Commission (1983). PRA Procedures Guide. A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants. Final Report. NUREG/CR-2300. Washington, DC: US Nuclear Regulatory Commission. http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/vol2/cr2300v2-a.pdf (accessed April 16, 2014).
von Mises, R. (1928) Wahrscheinlichkeit, Statistik und Wahrheit. 2nd edition, 1936. Vienna, Austria: Springer-Verlag.
Wakker, P.P. 2010 "Prospect theory for risk and ambiguity" Cambridge U. Press, Cambridge, 2010.)
Walley, P. (1991) Statistical Reasoning with Imprecise Probabilities. London, UK: Chapman and Hall.
Weitzman, M.L. (2001) Gamma discounting. American Economic Review 91 (1), 260–271.
Wielicki BA et al (2013) Achieving climate change absolute accuracy in orbit. Bull Am Meteorol Soc 94:1519– 1539. https://doi.org/10.1175/BAMS-D-12-00149.1
Wielicki BA, Barkstrom BR, Harrison EF, Lee RB III, Smith GL, Cooper JE (1996) Clouds and the Earth's Radiant Energy System (CERES): an earth observing system experiment. Bull AmMeteorol Soc 77(5):853– 868
Zelinka MD, Randall DA, Webb MJ, Klein SA (2017) Clearing clouds of uncertainty. Nat Clim Chang 7:674– 678
Zhou C, Zelinka MD, Dessler AE, Klein SA (2015) The relationship between interannual and long-term cloud feedbacks. Geophys Res Lett 42:10,463–10,469. https://doi.org/10.1002/2015GL066698
Zhou C, Zelinka MD, Klein SA (2016) Impact of decadal cloud variations on Earth's energy budget. Nat Geosc 9:871–874. https://doi.org/10.1038/ngeo2828