

Deep and shallow uncertainty in messaging climate change

Roger M. Cooke

Resources for the Future, Department of Mathematics, Washington DC and Delft University of Technology, Delft, The Netherlands

ABSTRACT: Deep and shallow uncertainty are defined and contrasted with regard to messaging the uncertainty about climate change. Deep uncertainty is often traced back to the writings of Frank Knight, where in fact it simply meant subjective probability. Although Knight envisioned a scientifically grounded quantification of subjective uncertainty, deep uncertainty is frequently invoked to disable quantification, with attendant problems in communicating and propagating uncertainty through chains of reasoning. These issues together with science based uncertainty quantification are illustrated with recent applications to ice sheet dynamics. The issues of performance assessment and validation are addressed.

1 THE INEVITABILITY OF UNCERTAINTY QUANTIFICATION

The US National Research Council 2010 report *Advancing the Science of Climate Change* (http://www.nap.edu/catalog.php?record_id=12782) illustrates reasoning and communicating under uncertainty. Using the IPCC AR4 calibrated uncertainty language, the report bases its first summary conclusion on *high confidence* (at least 8 out of 10) or *very high confidence* (at least 9 out of ten) in six statements¹ (p. 4,5):

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.

What is the confidence that all six of these statements hold? Pick a random integer between 1 and 10. You can have *high confidence* that it is strictly

less than 9 and also *high confidence* that it is strictly greater than 2. Do you have *high confidence* that it is one of the six integers 3, 4 ... 8? If six statements each have a 0.8 chance of being true, the chance that all are true could be anything from 0 to 0.8. Note that in the natural language 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 is a very poor vehicle for reasoning under uncertainty.

It gets worse. Consider the second statement. Are the authors highly confident that ‘the earth is warming AND humans are responsible’, or are they highly confident that ‘GIVEN that the earth is warming, humans are responsible’? These are very different statements. Since the Earth’s warming is asserted in the first statement, perhaps the second statement is meant. In that case, the likelihood of both statements holding is the product of their individual likelihoods. If the first two statements enjoy “high confidence”, then both can hold with only “medium confidence”. Scouring the text for the authors’ meaning is futile; were there any determinate meaning, no scouring would be needed. Instead of reasoning under uncertainty, the authors do what most of us do most of the time, they just throw words at it.

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 by the seat of your pants, as the National Research Council has amply shown. We are now facing decisions

1. “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.”

that could eventually impact the habitability of the planet as we know it, and we will take those decisions without knowing exactly how current actions affect the future climate. It's all about uncertainty and, not surprisingly, uncertainty has become a key part of the climate messaging: deniers and contrarians use uncertainty to shift the proof burden, alarmists use uncertainty to frighten us into action; science messengers, crafty or clumsy, magnify their certainty. The way forward starts with getting the uncertainty narrative right, and this involves more than hurling words.

Specialists in topics newly confronting decision under uncertainty often improvise their own approaches. Following the introduction of expert systems in the 1970's, the artificial intelligence community experienced an explosion of "alternative representations of uncertainty" through the 1980's, including certainty factors, degrees of possibility, fuzzy sets, belief functions, random sets, imprecise probabilities, non-monotonic logic, imprecise probabilities, among many others. The proceedings of the premier conference Uncertainty in Artificial Intelligence have been digitized since 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" accounts for 26% of the total.

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

Alternative representations of uncertainty re-emerge in new fields. This paper is a high pass over the their entrance in climate science, reducing technical jargon to the minimum. The logic of partial belief is rehearsed at the 30,000 ft level. Imprecise

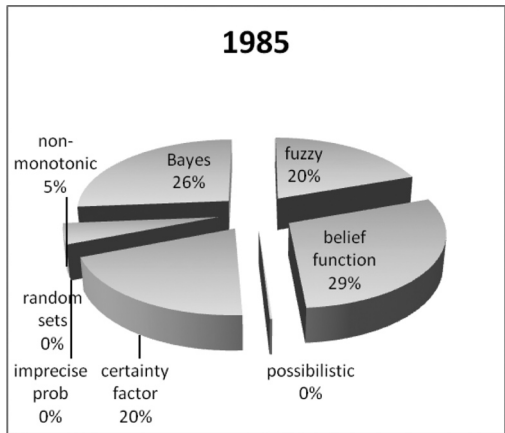


Figure 1. Word fragment counts in Uncertainty in Artificial Intelligence, 1985.

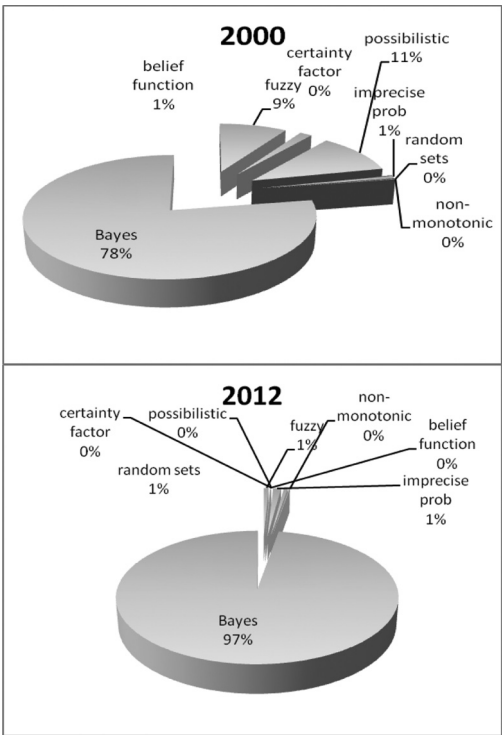


Figure 2. Word fragment counts in Uncertainty in Artificial Intelligence, 2000 (left) and 2012 (right).

probabilities, Deep, and Knightian uncertainties are overflowed. For some uncertainties, the apparent depth may result from the lack of operational meaning; this is called shallow uncertainty. The alternative to hurling words at uncertainty is to quantify it through a process of rational consensus based on empirical measures of performance. The last two sections introduce this.

2 THE LOGIC OF PARTIAL BELIEF

The subjectivist interpretation of probability goes back to Thomas Bayes (1763), but the operational measurement of partial belief as subjective probability dates from Ramsey (1926). 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 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 relation is additive:

\$10,000 if France OR Belgium ...; \$100 otherwise
 is preferred to
\$10,000 if USA OR Belgium ...; \$100 otherwise.

Other axioms require that the qualitative probability relation is transitive, that it is not influenced by the size of the (positive) rewards 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 more probable than 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. Suppose we partition all possibilities into disjoint events $A_1 \dots A_{2n}$ arranged in increasing qualitative probability, and such A_{2n} is not qualitatively more probable than $A_1 \cup A_2$. Form a new partition $B_1 \dots B_n$ such that $B_1 = A_1 \cup A_{2n}$, $B_2 = A_2 \cup A_{2n-1}$, etc. The B 's will be more uniform than the A 's. Suppose that event X is qualitatively more probable than the union of the first k of the B '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, 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. 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".*²

The logic of partial belief is similar; it does not tell us the degree to which we should believe that "Bill, who is dull and good in math, is an accountant", but it says the degree to which we believe "Bill is a jazz musician" cannot be smaller than the degree to which we believe "Bill is a jazz musician and an accountant" (Tversky and Kahneman

1982). Yet people commit such errors in droves. The paradoxes of Allais (1953) and Ellsberg (1961) describe choice situations in which many people exhibit behaviour 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 such paradoxical choice behaviour 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 empirically observed behaviour patterns.

To be sure, there are limitations in the representation of partial belief as subjective probability:

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: *\$10 if I clean my cellar next week, \$0 else* would not be preferred to *\$10 if Heads with a fair coin, \$0 else*; but changing the rewards from \$10 to \$1,000,000 in both lotteries would change my preferences.
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.

3 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

2. 80% of third year mathematics students at the Delft Technical University 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".

example Fu et al. 2005, Hall et al. 2007, Ghosh and Mujumdar 2009, Krieglert et al. 2009). Without disputing the substantive contribution these authors have made, we briefly discuss the use imprecise probabilities in the 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 exceeds 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, ... order imprecisions can deal with that. The uninitiated ask, “if you’re uncertain about a value between 0.1 and 0.5, 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 explained as follows:

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. (Ferson et al. 2007, p. 19).

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 o(US Nuclear Regulatory Commission (1983) 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. (p. 12–12)

If different probability intervals are generated by different experts, what do we do with them? Following Hall et al. (2007) we could take weighted combinations of the lower bounds, in order to be “conservative”. If there is no way to distinguish good and poor probability interval assessments then we can do anything else with equal justice.

A very large measure of arbitrariness is introduced in this way.

Fuzziness as a representation of uncertainty has also appeared in the context of emissions scenarios (Fu et al. 2005). Regarding the question whether fuzziness represents uncertainty, the discussion can be very brief. Suppose you get an email from an unknown Quincy, and you are equally uncertain that 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 $[0,1]$ reflecting the degree to which Quincy is believed to be in the set of Men. A similar function $\mu_W(Q)$ describes the degree to which Quincy is believed to be in the set of Women. 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 Men and Women. 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, $\mu_M(Q) = \mu_W(Q) = 1/2$; and Quincy is a Man AND a Woman with value $1/2$. Your uncertainty that Quincy is either a Man or a Woman would be the maximum of these uncertainties, also $1/2$. Combination rules have proliferated, but they all share the 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 does not behave that way.

4 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 originate in 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).” U.S. Senate Committee on Commerce, Science and Transportation Hearing on “The Case for Climate Change Action” October 1, 2003 Stephen H. Schneider.

A search for a precise definition came no farther 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, into which breach the imprecisionists have sprung. The sobriquet is that deep uncertainty is “Knightian”.

Economist Frank Knight’s book *Risk, Uncertainty and Profit* (1921) appeared in the same year as John Maynard Keynes’ *Treatise on Probability*, well before F.P. Ramsey’s operational definition of partial belief in terms of observed preference behaviour (1931) and before R. von Mises’ frequentist interpretation of probability (1928). 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. For Knight, “risk proper” is measurable by resolving outcomes into equi-probable alternatives (Knight 1921, III.VII.34). By aggregating risks, losses can be converted 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 Mormon 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.”

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:

“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 1921, III.VIII.1).

Knight, writing in 1921, did not know how to measure subjective probabilities. Neither did he know how to measure “risk” or objective probabilities. 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, were not well understood.

This is not to gainsay that Knight was ahead of his times. The idea of calibrating the probability judgments of the individual “business man” prefigures the modern use of structured expert judgment.

“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 ...” (Knight, 1921 III.Vii.43)

5 SHALLOW UNCERTAINTY

Shallow uncertainty is uncertainty resulting from undefined terms, careless formulation, lack of operational definitions 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.

As an example, consider the Social Discount Rate (SDR) often written as:

$$\text{SDR} = \rho + \eta G(t)$$

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 . It is generally recognized that the discount rate is an important driver, if not the most important driver, in the economic models for climate change. 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” (p. 60) or “taste variables” (p. 215) which 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, Lowenstein and O’Donoghue (2002, p. 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 discounted utility, 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]” (p. 355). Weitzman (2001) and Pizer and Newell (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). It is not appropriate for (a) if we are the chooser of policies. Various authors assign time preference and risk aversion to both (a) and (b).

Questions arise as to the operational meaning of society’s rate of pure time preference. 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 modelling community has not agreed on operational definitions for such terms, then the models are not ready for uncertainty analysis. The adage is: If you don’t know what it means, you can’t be uncertain about it.

6 EXPERTS

Saying ‘we don’t know the probability distribution’ implies that probability is an objective property of the world which we happen not to know. Partial belief is not a property of the world in this sense. Introspection suffices for quantifying partial belief; there is nothing to not know. The rub is that each individual is entitled to his/her own subjective probability distributions. Those most knowledgeable about a scientific field, experts, will also have differing degrees of partial belief about variables of interest. Expert disagreement is inherent in choice under uncertainty. Indeed, if the science ‘isn’t there yet’, experts are supposed to disagree. This section

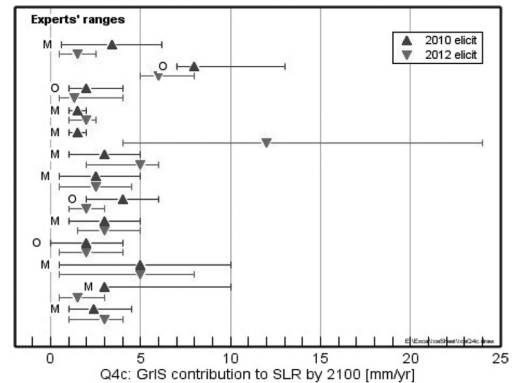


Figure 3. 21st century contribution to sea level rise from the Greenland ice sheet [mm/yr].

broaches the issue of using experts to quantify uncertainty. If shallow uncertainty regarding the meaning of words is under control, there is no impediment to experts quantifying their degrees of belief as subjective probability distributions.

Data is helpful at this point. The following graph from Bamber and Aspinall (2013) shows the results of eliciting subjective probability distributions from 12 experts regarding the 21th century contribution to sea level rise from the Greenland ice sheet in mm/yr. The same experts were elicited in 2010 and in 2012. “M” means that the expert is primarily a modeller, “O” denotes an observationalist. Experts assessed the 5, 50 and 95 percentiles of their subjective probability distributions. A value of 3 mm/yr means that by 2100 the Greenland ice sheet will have contributed roughly 15 cm to sea level rise.³

The experts’ 95 percentiles for this contribution range from 15 cm to 120 cm for the 2012 elicitation. What are we to do with this? Clearly we could cherry pick different experts and convey very different messages. Some authors argue that the experts’ distributions should not be combined; the diversity of expert views is the proper message to the decision makers (Morgan et al. 2009, Morgan and Henrion 1990). Indeed, if the raw uncombined expert data provide sufficient information to take a decision, then combination is unnecessary. In many cases, like the Greenland ice sheet example, the raw data is not sufficient; if we are 95% sure that this contribution is less than 15 cm in 2100, we would take different decisions than if there is a 1 in 20 chance of a contribution exceeding 1.2 meters.

A strong argument in favour of non-combination is that any combination procedure requires

3. This assumes effectively zero contribution in 2000 and linear growth rate of the contribution to its 2100 value.

justification in order to deflect allegations of cherry picking. Combining the experts' distributions with equal weight entails justifying that experts perform these probabilistic assessment tasks equally well. What does it mean to perform probabilistic assessment tasks well or badly? A discussion of performance must precede a discussion of combination. Expert performance data was collected for the ice sheet study, but not published⁴; we turn to other sources.

6.1 Expert performance

An expert assessing an unknown quantity can be regarded as a new sort of measuring instrument. The scientific employment of this instrument requires trying it out on things we know before applying it to things we don't know. With sufficient time and patience we could ask experts to give point predictions of things we later observe, and assess each experts' error distribution. It doesn't sound easy, but it sounds easier than it is. How do we add the error distance for an item expressed in nanometres with another expressed in kilograms per square kilometer? If we change the first to meters, should we change the second? Even if the variables are both in years, an error of 3 years in assessing the mean age of an aircraft fleet is not the same as an error of 3 years in assessing the age of your youngest daughter. To assess performance, the variables must be converted to a common scale, and the common scale is probability. Instead of an error distance between an estimated and observed quantity, performance will focus on 'how surprised' the expert should be by the observed quantity. That requires that experts assess their subjective distributions for the unknown quantities. Experts then become statistical hypotheses, amenable to empirical control via standard statistical methods.

A recent study of information security (Ryan et al. 2012) used 10 calibration questions to score performance. Figure 2 shows the 5, 50, and 95 percentiles of the best and worst performing experts in a panel of 13, and also shows the realizations (#). The calibration questions concerned attack rates on different information systems. Expert 1 catches one of the 10 realizations in his 90% central confidence band, whereas expert 2 catches 9 out of 10. Should we use such information, and if so, how?

6.2 Bayesian approaches

Most statisticians, confronted with this sort of data for the first time, would seek to combine the

experts' performance information through some form of Bayesian updating. There have been many proposals in this direction, starting with Morris (1977), yet they have come to naught. With a bit of formalism, it is easy to understand why (Cooke 1991).

Suppose we have two unknown continuous quantities X_1 and X_2 . One expert gives his cumulative distribution functions F_1, F_2 for X_1 and X_2 . The decision maker also has prior distributions G_1, G_2 for X_1 and X_2 . The decision maker wants to use the observed value of X_1 , say x_1 , to update his distribution about X_2 . The decision maker has a joint distribution for X_1 and X_2 , and (s)he should simply update that whereby the information from the expert plays no role. However, we suppose that somehow the decision maker can't do that, and instead wants to update his opinion about this expert based on the observation of x_1 . For any r in $(0, 100)$, $\{F_1(X_1) \leq r/100\}$ denotes the event that X_1 takes a value less than or equal to the r -th percentile of the expert's distribution. The expert evidently believes that this event has probability $r/100$, but the decision maker may assign it a different probability. If the decision maker thinks this expert is likely to underestimate X_1 , then his probability of $\{F_1(X_1) \leq r/100\}$ will be less than $r/100$. The decision maker will update his probability of event $\{F_2(X_2) \leq r/100\}$ based on observing $\{F_1(X_1) \leq r/100\}$. This requires specifying the decision maker's joint probability of these events. However, if he does that, problems will arise. The simplest case is that these two events are *exchangeable*, for any r . This entails that for the decision maker, these two events are equally likely, for any r . Now we do a simple calculation, where P denotes the decision maker's probability, and assuming the distribution functions are all invertible:

$$\begin{aligned} P(\{F_1(X_1) \leq r/100\}) &= P\{X_1 \leq F_1^{-1}(r/100)\} \\ &= G_1 F_1^{-1}(r/100) = P(\{F_2(X_2) \leq r/100\}) \\ &= G_2 F_2^{-1}(r/100); \end{aligned}$$

From which we infer, for any r

$$F_2^{-1}(r/100) = G_2^{-1} G_1 F_1^{-1}(r/100).$$

The expert's distribution for X_2 is completely determined by his distribution for X_1 and the decision maker's priors. The decision maker needn't consult experts at all!

6.3 Rational consensus

The Bayesian paradigm is a paradigm for *individual* decision making. An expert judgment study can have any of several possible goals: an expert survey,

4. Publications with performance data are in preparation at this writing.

a political consensus seeking an equilibrium of stakeholders, or a rational consensus. Rational consensus is not formally defined, but the idea is this: for combining their opinions, experts pre-commit to a method of combination that satisfies necessary conditions of the scientific method. Then, the method is executed. Experts needn't adopt the results as their personal probability distributions, but withdrawing from the rational consensus incurs a proof burden. They must argue that other necessary conditions for the scientific method have not been met. Withdrawing for other reasons, for example, that the results are hostile to one's interests, does not threaten rational consensus.

For details on combining experts see (Cooke 1991, Cooke and Goossens 2008, Aspinall and Cooke 2013). A few points can be recalled here. The most important requirement of the scientific method is empirical control. This requires that each expert-hypothesis be empirically tested on the basis of calibration variables from the experts' field whose values are known post hoc. Figure 2 is an example. For each item, each expert's percentiles (5, 50, 95) define four intervals to which (s)he assigns realization probabilities 0.05, 0.45, 0.45, 0.05. The statistical hypothesis being tested is that the realizations can be viewed as independent samples from this distribution.⁵ The p-value of each hypothesis, given the realizations, is the probability of falsely rejecting the hypothesis that the expert's probability statements accord with the realizations. Note the eight orders of magnitude difference in these p-values. Statistical likelihood is a fast function. If you believe that 10 independent events each have probability 90% of occurring, then the likelihood of seeing only one of the ten occur is $10 \times 0.9 \times 0.1^9 \sim 10^{-8}$ (the computation for the data in Figure 2 is a bit more complicated, owing to the presence of the 50 percentile, and also a chi square approximation). Each expert's statistical accuracy

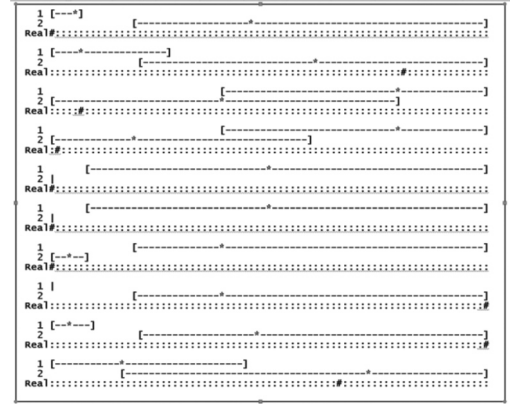


Figure 4. Best and worst performing experts in Ryan et al. (2012).

is measured by the corresponding p-value. Less important than statistical accuracy, but still important, is the expert's informativeness. Intuitively, this is measured as the expert's ability to concentrate high probability mass in small intervals. The measure of informativeness must not depend on the scale of the underlying variables. Cognoscenti wanting details are referred to the above references. These performance measures are combined so as to satisfy a strictly proper scoring rule constraint: in an appropriate long run sense, an expert maximizes his/her expected score by and only by stating percentiles corresponding to his/her true beliefs. It must be emphasized that measuring expert performance and combining experts satisfies constraints devolving from the scientific method but are in no way unique. In the same way, a bicycle satisfies Newton's laws but does not follow from these laws; it is designed to serve a purpose. In the same way the performance-based combinations of expert judgments are designed to serve the purpose of finding rational consensus on the quantification of uncertainty. Better bicycles are always welcome.

7 EXPERT JUDGMENT VALIDATION

Forty five professionally contracted studies with domain experts from fields including of nuclear safety, aerospace and aviation risk, environmental transport, finance, volcanology, public health were reviewed by Cooke and Goossens (2008) in a special issue of *Reliability Engineering and System Safety*. The data including calibration variables was made available to researchers. Highlights of research based on this data are summarized below. Since then the applications have nearly doubled, expanding into areas like ecosystem modelling, project management, and climate modelling. All

5. In this context, independence is not an assumption about the experts' joint distribution, it is a desideratum of the decision maker. If the experts' joint distributions are dependent, then the expert will learn from observing the calibration variables. Expert learning is not a goal of the elicitation, and the decision maker desires experts who have already learned. If the calibration variables are multiple realizations of the same random variable, or otherwise confounded, then standard techniques for reducing the power of the statistical test apply.

studies compared an Equal Weight (EW) combination with performance based weighted combination (PW). Measuring performance transforms the field of expert judgment from a confessional to a science- and data-based activity. Any combination of experts' distributions can be tested against the set of calibration variables, and its performance scored. When scored on the calibration variables, equal weighting usually—not always—produces statistically acceptable performance at the expense of low informativeness. Performance-based combinations usually—but not always—perform better on both counts. Both PW and EW may be scored for statistical accuracy and informativeness; combining scores for statistical accuracy and informativeness yields a “combined score” (Cooke and Goossens 2008).

In a few studies, variables of interest were later observed, enabling true out-of-sample validation (Cooke and Goossens 2008). In most cases the variables of interest are not observable on timescales relevant for the decision problem. Therefore, various forms of cross validation have been suggested. Clemen (2008) proposed a Remove-One-At-a-Time (ROAT) method according to which the calibration variables were removed one at a time and predicted by the model initialized on the remaining calibration variables. The predictions, though originating from different decision makers, were pooled and compared with the equal weight decision maker. On the 14 studies selected for this exercise, Clemen found that PW combined score was higher than that of EW on 9, which was not statistically significant. Cooke (2008a, 2012) noted that this procedure is biased against PW since each calibration variable is predicted by a decision maker in which experts who assessed that particular item badly are up-weighted. It is commonly observed that removing one calibration variable can influence an individual expert's statistical likelihood by a factor 3 or more, a feature explained by the fact that statistical accuracy is a very fast function.

Variations on the ROAT approach have been performed by other researchers. Lin and Cheng (2008) examined 28 of the 45 studies and found PW significantly out performing EW, although PW's out-of-sample performance was degraded. Lin and Cheng (2009) used ROAT on 40 studies finding no significant difference between PW and EW.⁶ Lin and Huang (2012) used ROAT with the Brier score in a regression based study of the effects of aggregation method, dependence, number of experts and seed variables and overconfidence on the Brier score.

6. There large differences between the in-sample values in these two papers, and those found in the original studies.

Other researchers have undertaken cross validation without ROAT. Cooke (2008a) looked at half-half splits in 13 studies with at least 14 calibration variables. Flandoli et al. (2010) examined five datasets, choosing 30% of the number of calibration variables as the size of the test set, provided this number was at least 8, otherwise the test set was 8. They recoded the classical model in R with some restrictions⁷ and randomly drew 500 partitions into training and test sets of the fixed sizes. The most extensive study of this kind is Eggstaff et al. (2013), which initializes the global weights model on *all* non empty subsets of seed variables and in each case predicts the complementary subset, again using only global weights. Studies with large numbers of seed variables were split into separate studies to prevent combinatoric explosion. In total 62 expert judgment studies were analysed.

Studies differ in subject matter, number and training of experts, the methods of recruitment and methods of elicitation. For this reason, a numerical representation of out-of-sample validity at the study level would be desirable. For N seed variables, for each $K = 1 \dots N-1$, Eggstaff et al. (2013) compute the combined scores of PW and EW for each K -tuple of variables in the training set, based on prediction of $N-K$ out-of-sample variables in the test set. The same experts, the same calibration variables, and the same information background measures apply for all training set choices within one study. However the statistical power of the test set goes down as the training set size increases, there are many more K -tuples for values of K near $N/2$, and these studies have overlapping training sets. With this in mind the PW and EW combined scores are averaged for each size K , for $K = 1 \dots N-1$. The ratio of PW and EW can be compared across training set sizes. In aggregating ratios of positive numbers we must take the geometric mean, or geomean.⁸

7. Specifically, they did not implement the log uniform background measure, and did not implement “item weights” which are item specific and depend on the experts' informativeness for each item. This is the performance weighting scheme most often used in practice.

8. To see this suppose on two comparisons the scores were (PW = 4, EW = 1) and (PW = 1, EW = 4) The performance is identical, but the average of ratios is $1/2(4 + 1/4) = 2.125$. The Geomean is $(4 \times 1/4)^{1/2} = 1$. Eggstaff et al. report only the average scores for each size of the training sets, so we consider the ratios of averages. Since the average is always greater or equal to the geomean, the numerator and denominator in these comparisons would both be smaller if we took the geomeans of combined scores of each separate K -tuple of training variables. It's impossible to say if there is an overall effect of this choice.

The geomean of the ratios of combined scores of all comparisons for each of the 62 studies are plotted in Figure 5. In 45 of the 62 studies (73%) the geomean of combined score ratios PW/EW was greater than unity. When PW's combined score exceeded that of EW, it tended to exceed by a greater amount than when EW's combined score exceeded that of PW. The best eyeball assessment is to compare the mass of lines above and below the baseline of 1. The geomean of the geomeans for each study was 2.46. Summarizing, PW outperforms EW in out of sample cross validation on more than two thirds of the studies, and the combined score of PW is more than twice that of EW.

The accuracy of a DM in terms of proximity of the median to the true value is not directly related to the scoring variables of statistical accuracy and informativeness. Eggstaff et al. (2013) report an accuracy advantage of PW over EW comparable to the ratios of combined scores; however that feature is not pursued in this paper.

The extensive dataset of domain expert elicitations has afforded other insights into expert judgment. Based on a number of studies in which experts were very well published, it was possible to derive weights based on experts' citation indices (Cooke, Elsaadny and Huang 2008) . Performance was comparable to equal weighting and inferior to performance-based weights. On a few occasions the analyst asked experts to rank each other, and compared these ranks with their performance on calibration variables. The findings for elicitations on ice sheets and dam safety in Figure 6 speak for themselves:

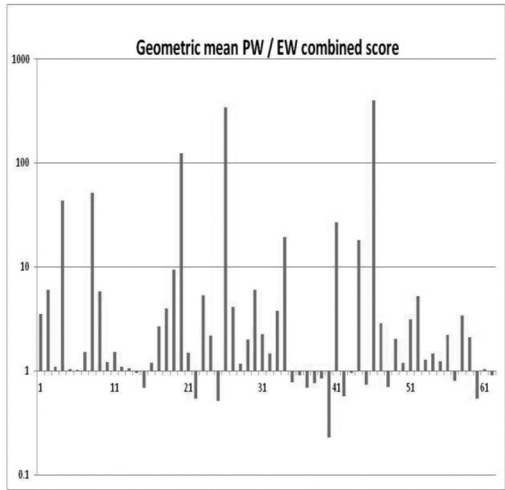


Figure 5. 62 studies, per study: geomeans of comparisons of PW/EW combined score ratios.

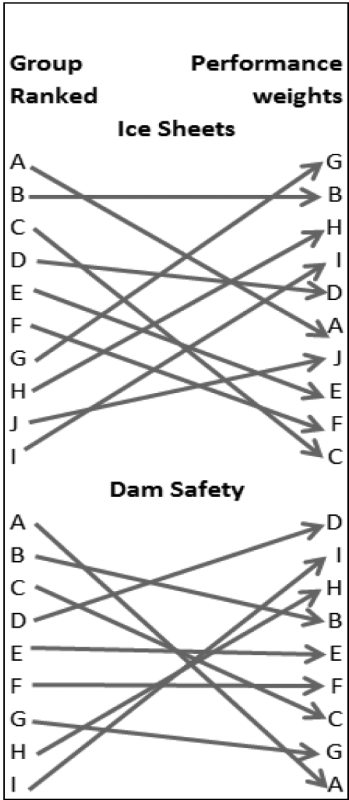


Figure 6. Expert mutual self-rankings and performance ranking, Ice sheets (Nov. 2012) left and Dam safety right (Aspinall and Cooke, 2013).

This should not come as a total surprise. Little in an expert's training prepares him or her for probabilistic assessment. Some experts are able to give very accurate and informative assessments, many others less able to do this. Unfamiliarity with probabilistic assessment will be reflected not only in expert scores but in experts' ranking of colleagues for probabilistic assessment tasks. This simple observation has noteworthy consequences. Expert nominations for advisory roles, blue ribbon panels and the like may identify experts with substantive knowledge, influence and poids in their field, but the supposition that this is a good way to quantify uncertainty is not supported by the evidence.

Returning to the Greenland ice sheet, Figure 7 shows combinations based on self weights, equal weights and performance weights, for both elicitations, in 2010 and 2012 respectively.

From the more recent elicitation, the equal weights decision maker's 95 percentile for total sea level rise in 2100 due to the Greenland ice sheet is about 80 cm, and for self weights about 60 cm; for

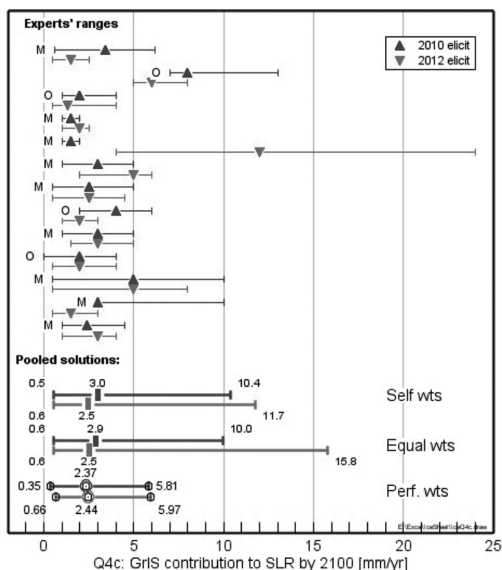


Figure 7. Greenland ice sheet contribution to sea level rise in 2100, with combined experts.

the performance-based decision maker the value is close to 30 cm.

7 CONCLUSIONS

Whether the calibrated language, deep uncertainty, Knightian uncertainty, imprecise probabilities in climate change will go the way of fuzzy sets, random sets, belief functions, and degrees of possibility in artificial intelligence remains to be seen. However useful the alternative representations of uncertainty may be, they should not crowd out active research within the probabilistic approach. In addition to expert validation discussed above, other areas include, above all, high dimensional dependence modelling. After the normal copula crashed and burned in 2008 (Salmon 2009), financial mathematicians have engaged actively in capturing dependence in the tails of random variables. Complex graphical structures are emerging for characterizing and computing with these dependence structures. Dependence elicitation began in earnest in the nuclear sector 20 years ago, but has had limited application in other studies. However, very recent work has introduced the concept into assessing future ice sheet melting due to global warming. Once we have good measures of performance, training experts for comprehensive probabilistic assessment of complex problems becomes a possibility, the need for which has some urgency. On real applications there is seldom time

and budget for this, and little research has been done. Another cluster of questions concerns fitting models to expert judgment, and evaluating models on the basis of expert judgment.

REFERENCES

- Allais, M. (1953). Le comportement de l'homme rationnel devant le Risque. *Econometrica* 21, 503–546.
- Aspinall, W. (2010). “A route to more tractable expert advice” *Nature*, vol. 463, 21 January.
- Aspinall, W.P. & Cooke, R.M. (2012). 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 (in press).
- Clemen, R.T. (2008). “Comment on Cooke’s classical method” *Reliability Engineering & System Safety*, vol. 93, no. 5, May 2008, pp. 760–765.
- Cooke R.M. (1991). *Experts in Uncertainty; Opinion and Subjective Probability in Science*, Oxford University Press; New York Oxford.
- Cooke, R.M. (2004). “The anatomy of the Squizzle—the role of operational definitions in science” *Reliability Engineering and System Safety*, 85, 2004, 313–319.
- Cooke, R.M. (2012). “Pitfalls of ROAT Cross Validation Comment on Effects of Overconfidence and Dependence on Aggregated Probability Judgments” appearing *Journal of Modelling in Management*, vol. 7.
- Cooke, R.M. (2003). Book Review Elicitation of expert opinions for uncertainty and risks Elsevier, *Fuzzy Sets and Systems* 133, pp. 267–268, ISBN 0-8493-1087-3.
- Cooke, R.M., ElSaadany, S. & Huang, X. (2008). On the Performance of Social Network and Likelihood Based Expert Weighting Schemes, Special issue on expert judgment *Reliability Engineering & System Safety*, 93, 745–756, Available online 12 March 2007, vol. 93, no. 5, May 2008.
- Cooke, R.M. & Goossens, L.H.J. (2008). TU Delft Expert Judgment Data Base, Special issue on expert judgment *Reliability Engineering & System Safety*, vol. 93, no. 5, pp. 657–674.
- Cooke, R.M. (2008a). Response to Comments, Special issue on expert judgment *Reliability Engineering & System Safety*, 93, 775–777, Available online 12 March 2007, vol. 93, no. 5, May 2008.
- Eggstaff, J.W., Mazzuchi, T.A. & Sarkani, S. (2013). The Effect of the Number of Seed Variables on the Performance of Cooke’s Classical Model, appearing in *Reliability Engineering and System Safety*.
- Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics* 75, 643–669. (DOI: 10.2307/1884324).
- Evans, D.J. & Sezer, H. (2005). Social discount rates for member countries of the European Union, *Journal of Economic Studies*, vol. 32, no. 1, pp. 47–59.
- Ferson, S. Kreinovich, V. Hajagos, Oberkampf, S. & Ginzburg, L. (2007). *Experimental uncertainty estimation and statistics for data having interval uncertainty*. SAND2007–0939, prepared by Sandia National Laboratories, Albuquerque, NM and Livermore, CA.
- Flandoli, F., Giorgi, E., Aspinall, W.P. & Neri, A. (2010). “Comparison of a new expert elicitation model

- with the Classical Model, equal weights and single experts, using a cross-validation technique". *Reliability Engineering and System Safety*, 96, 1292–1310. doi:10.1016/j.ress.2011.05.012.
- Fredrick, S., Loewenstein, G. & O'Donoghue, T. (2002) "Time discounting and time preference: a critical review" *Journal of Economic Literature* vol. 40, no. 2, pp. 351–401.
- Fu, G., Hall, J. & Lawry, J. (2005). "Beyond probability: new methods for representing uncertainty in projections of future climate" Tyndall Centre for Climate Change Research Working Paper 75.
- Ghosh, S. & Mujumdar, P.P. (2009). "Climate change impact assessment: Uncertainty modeling with imprecise probability" *JOURNAL OF GEOPHYSICAL RESEARCH*, VOL. 114, D18113, doi:10.1029/2008 JD011648, 2009.
- Hall, J.W., Fu, G. & Lawry, J. (2007). "Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties" *Climatic Change* April 2007, Volume 81, Issue 3–4, pp 265–281.
- Henry, C. (2006). "Decision-Making Under Scientific, Political and Economic Uncertainty." *Laboratoire d'Econométrie de l'Ecole Polytechnique Chair Développement Durable Cahier DDX–06–12*.
- IPCC (2005). Guidance notes for lead authors of the IPCC Fourth Assessment Report on addressing uncertainties. Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland. <http://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf>.
- Kandlikar, M., Risbey, J. & Dessai, S. (2005). "Representing and communicating deep uncertainty in climate-change assessments", *C.R. Geoscience* 337 (2005) 443–455, doi:10.1016/j.crte.2004.10.010.
- Knight, F.H. (1921). *Risk, Uncertainty and Profit*. New York: Harper & Row. <http://www.econlib.org/library/Knight/knRUPCover.html>.
- Krieger, E., Hall, J.W., Held, H., Dawson, R. & Schellnhuber, H.J. (2009). "Imprecise probability assessment of tipping points in the climate system" *PNAS*, March 31, 2009, vol. 106, no. 13, pp. 5041–5046.
- Lenton, T.M., Held, H., Krieger, E., Hall, J.W., Lucht, W., Rahmstorf, S. & Schellnhuber, H.-J. (2008). "Tipping elements in the Earth's climate system", *PNAS* vol. 105, no. 6, pp. 1786–1793.
- Lewis, H. et al. (1979). *Risk Assessment Review Group Report to the US Nuclear Regulatory Commission NUREG/CR-04000*.
- Lin, Shi-Woei & Bier, V.M. (2008). "A Study of Expert Overconfidence" *Reliability Engineering & System Safety*, 93, 775–777, Available online 12 March 2007. vol. 93, no. 5.
- Lin, Shi-Woei, Cheng & Chih-Hsing (2008). "Can Cooke's Model Sift Out Better Experts and Produce Well-Calibrated Aggregated Probabilities?" Department of Business Administration, Yuan Ze University, Chung-Li, Taiwan Proceedings of the 2008 IEEE IEEM.
- Lin, Shi-Woei, Cheng & Chih-Hsing (2009). "The reliability of aggregated probability judgments obtained through Cooke's classical model", *Journal of Modeling in Management*, vol. 4, no. 2, pp. 149–161.
- Lin, Shi-Woei, Huang & Ssu-Wei. (2012). "Effects of Overconfidence and Dependence on Aggregated Probability Judgments" *JMM*, appearing March 2012.
- Mach, P.R., Matschoss, G.-K., Plattner, G.W., Yohe & F.W. Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). <http://www.ipcc.ch> (2010).
- Mastrandrea, M., Mach, K., Plattner, G.-K., Edenhofer, O., Stocker, T., Field, C., Ebi, K. & Matschoss, P. "The IPCC AR5 Guidance Note on Consistent Treatment of Uncertainties: A Common Approach across the Working Groups", *Climatic Change* DOI 10.1007/s10584-011-0178-6 (2011). <http://www.climate.unibe.ch/~stocker/papers/mastrandrea11cc.pdf>.
- McCrimmon K.R. (1968). Descriptive and normative implications of the decision-theory postulates. In: *Risk and Uncertainty*. MacMillan, London pp. 3–24.
- Morgan, G.M., Dowlatabadi, H., Henrion, M., Keith, D., Lempert, R., McBride, S., Small, M. & Wilbanks T. (2009). "Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decisions", U.S. Climate Change Science Program Synthesis and Assessment Product 5.2.
- Morgan, M. & Henrion M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge University Press, New York.
- Morgan, M.G. & Keith, D.W. (1995). Subjective judgments by climate experts. *Environmental Science & Technology*, vol. 29, no. 10, pp. 468–476.
- Morgan, M.G. (2011). *Certainty, uncertainty, and climate change*, *Climatic Change* DOI 10.1007/s10584-011-0184-8.
- Morgan, M.G. & M. Henrion. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, UK, 348 pp. (see particularly Chapter 6, "Human judgment about and with uncertainty".)
- Morris, P. (1977). "Combining expert judgments: A Bayesian approach" *Management Science*, vol. 23, no. 7, pp. 679–693.
- Moss, R.H. & Schneider, S.H. (2000). Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In: Pachauri, R., Taniguchi, T., Tanaka, K. (eds) *Guidance papers on the cross cutting issues of the Third Assessment Report of the IPCC*. World Meteorological Organization, Geneva, pp. 33–51. <http://www.ipcc.ch/pdf/supporting-material/guidance-papers-3rd-assessment.pdf>.
- Moss, R.H. & 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* [eds. R. Pachauri, T. Taniguchi & K. Tanaka], World Meteorological Organization, Geneva, pp. 33–51.
- Newell, R.G. & Pizer, W.A. (2003). "Discounting the distant future: how much do uncertain rates increase valuations?" *J. of Environmental Economics and Management*, vol. 46, pp. 53–71.
- Nordhaus, W. & Popp, D. (1996). "What is the value of scientific knowledge: an application to global warming using the PRICE model" *Cowles Foundation discussion paper* no. 1117.
- Nordhaus, W.D. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven, CT, Yale University Press.

- Pizer, W.A. (1999). "The optimal choice of climate change policy in the presence of uncertainty" *Resource and energy economics*, vol. 21, pp. 255–287.
- Ramsey, F.P. (1926). "Truth and Probability", in Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*, Ch. VII, p. 156–198, edited by R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company.
- Salmon, F. (2009). "Recipe for disaster" *The formula that killed Wall Street* *Wired Magazine*, 17.03, 02.23.
- Stern, N. (2008). "The Economics of Climate Change" *American Economic Review*, vol. 98, no. 2, pp. 1–37.
- Thomas Bayes (1763), "*An essay towards solving a Problem in the Doctrine of Chances.*" Bayes' essay as published in the *Philosophical Transactions of the Royal Society of London*, vol. 53, p. 370, on Google Books.
- Tversky, A. & Kahneman, D. (1982). "Judgments of and by representativeness" in Kahneman, Slovic and Tversky *Judgment under uncertainty: heuristics and Biases*, Cambridge U. Press, Cambridge, 1982, pp. 84–98.
- U.S. NRC (U.S. 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. Rockville, MD. Available from: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/vol2/cr2300v2-a.pdf>.
- US Nuclear Regulatory Commission (1983). *PRA Procedures Guide*, NUREG/CR-2300, 1983.
- von Mises, R. (1928). *Wahrscheinlichkeit, Statistik und Wahrheit*. Wien 1928, Springer-Verlag, 2nd edition 1936.
- Wakker P.P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge Univ Pr.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London.
- Weitzman, Martin L. (2001). "Gamma Discounting" *Source: The American Economic Review*, vol. 91, no. 1 (Mar., 2001), pp. 260–271.