Chapter 3

Environmental Decision-Making under Uncertainty

Joe Roussos[†]

Richard Bradley[*]

Roman Frigg[*]

Abstract Extreme weather events like hurricanes occur rarely, but when they occur, they cause immense damage. How should decision-makers, both public and private, make decisions about such events? Such decisions face significant and often poorly understood uncertainty. We rework the so-called "confidence approach" to tackle decision-making under severe uncertainty with multiple models, and we illustrate the approach with the case study of insurance pricing using hurricane models. The confidence approach has important consequences for this case and offers a powerful framework for a wide class of problems.

Keywords Uncertainty • Confidence • Insurance Pricing • Averaging • Catastrophe Model • Hurricane Model • Extreme Weather Event • Climate Change

We thank Tom Philp for numerous discussions about hurricane modelling, and Jan-Willem Romeijn, Sean Gryb, Simon Dietz, and Jonathan Livengood for helpful discussions about different aspects of the paper. Thanks to Wenceslao J. Gonzalez for inviting us to present the paper at the conference *Philosophy of Science Today*, which took place at the Universidade da Coruña, Spain, in March 2020, and to the audience at the conference for stimulating questions.

† Institute for Futures Studies, Stockholm

* Department of Philosophy, Logic, and Scientific Method, London School of Economics

1. Hurricane Maria

When Hurricane Maria hit Dominica in September 2017 it devastated the island nation, causing landslides, widespread flooding, and damage to the roofs of almost every home. The prime minister, Roosevelt Skerrit, had to be rescued from his official residence.[1] The island lost all radio, cell phone and internet services after the storm.

Each year, summer in the Northern Hemisphere brings hurricanes like Maria to the west Atlantic. Also known as "tropical cyclones" or typhoons, these giant, rotating storms wreak havoc in the Caribbean and the south east of the USA, leading to deaths, evacuations, and billions of dollars of damage each year.

Maria was one of the worst storms on record, and part of the costliest hurricane season on record for the Atlantic, with the final bill for storm damage in 2017 exceeding £228bn. The root of the damage is the incredibly fast winds that hurricanes generate. The wind itself is strong enough to damage cars, trees, and houses. A "category 1" hurricane, the lowest grade, has winds up to 150 km/h, enough to snap branches off trees and cause flying debris. Category 3, the boundary for a "large" hurricane, has winds strong enough to rip the entire roof from a house. At category 4, not even the wooden walls on American houses withstand the storm. But most hurricane damage comes from water, whipped up by the roaring winds. Meteorologists call this "storm surge": sea water is pushed into fierce waves, metres higher than the usual sea-level. The water slams into the coast and causes flooding for kilometres inland. Coastal buildings are demolished by the frequent pounding of waves during the storm, and flooding seawater erodes beaches and coastal highways, and undermines the foundations of buildings. It is hard to typify the damage caused by category 5 storms like Maria, as this is a catch-all for wind speeds above 252 km/h. They are catastrophes almost without parallel.

2. Extreme Weather Events and Climate Change

Hurricane Maria is an example of an "extreme weather event". An extreme event is one that occurs relatively rarely but has huge impact when it occurs. Droughts, heavy rainfalls, floods, and heat waves, as well as the increased incidence of extremely high sea levels or the more frequent occurrence of particularly hot days are extreme weather events. Not only are these events highly destructive in themselves; they often also have devastating consequences. Heatwaves cause the deaths of vulnerable people; heavy rainfalls cause landslides; and the more frequent occurrence of hot days creates ideal conditions for wildfires.

In its most recent assessment report, the Intergovernmental Panel on Climate Change (IPCC), the United Nations body for assessing the science related to climate change, comes to the conclusion that climate change, whether driven by natural causes or human activities, can result in changes in the likelihoods of the occurrence or severity of extreme weather events.[2] The IPCC also reports that such changes have indeed been observed since about the 1950s. They report, for instance, that it is likely that the frequency of heat waves has increased in large parts of Europe, that anthropogenic greenhouse gas emissions are likely to be a contributing factor,

[1] See, for example, https://www.nytimes.com/2017/09/19/world/americas/hurricane-maria-caribbean.html

[2] IPCC (2013, 121). For detailed overview of expected changes in extreme weather events of different kinds and their likely causes see IPCC (2013, 110). For a discussion of climate change and the philosophical and methodological question that it raises see Bradley and Steele's (2015), Frigg, Thompson and Werndl's (2015b, 2015c), and Parker's (2018).

and that further changes in the future are very likely. They report similar findings for other extreme evens like extremely high sea levels and tropical cyclones.

If we don't want to be hit by these events unprepared, we have to plan. Individuals as well as organisations will have to adapt to the fact that in the future we are more likely to be exposed to extreme weather events than we were in the past, and that these events are likely to be more severe than the ones we have hitherto experienced. Adaption policies put measures into place to help people cope with the effects of climate change, for instance by building flood defences, making buildings hurricane-proof, or increasing the capacity of water storage facilities.

Adapting our infrastructure so that it is able cope with extreme events is expensive, and adaptation requires significant resources. But how much exactly should we invest into adaptation measures? This will depend on what we think the damage that such events cause will cost. On the one hand, we don't want to overspend and invest significantly more into adaptation than any potential damage would be. On the other hand, we don't want to underspend and make ourselves vulnerable to huge losses. To design concrete policies and to secure their implementation, an assessment of the potential damages and their associated costs is indispensable.

Nowhere else is the price of a disaster as "in our face" as in the insurance sector. An insurance company will put a price on the potential damage, assess the likelihood that the damage will occur, and then combine the two into the price of an insurance policy. If they price their policies too low, they will go bust; too high and no one will buy their products. Insurers have to strike the same kind of balance that adaptation policies do. In fact, many of the calculations are the same: how much a state should invest into, say, flood defences will depend on what the expected damage due to flooding is. Looking at insurance pricing therefore gives us insight into how such assessments are made, and the problems that they encounter.

3. Hurricane Insurance

For people living in Florida, or on a Caribbean island, the risk of hurricane damage to their home is one of the most serious they face. Naturally, an insurance industry has grown around this risk, offering home-owners protection against the various forms of destruction hurricanes can bring. Residents buy insurance policies that guard them against such damage, at frequently high cost: a house insured for £120,000 will cost £2,500-6,000 *per year*.[3]

The price is so high in part because hurricanes cause significant damage, but also because insurers are so *uncertain* about how risky it is to insure. If you want to sell insurance against something the recipe is simple, with just three ingredients. First, you need the likelihood of the event you're covering (the hurricane). Second, you need an estimate of how damaging these events are when they occur—how much damage, in pounds, does the average hurricane cause in a £120,000 house? Third, you need to obey insurance regulations that tell you how much money you need to have available at any given time. These rules exist to ensure that insurance companies don't go bankrupt and have the money to pay for claims when customers make them.

But the first two ingredients are difficult to work out for hurricanes. Calculating the likelihood of destructive hurricanes requires a detailed understanding of the science of meteorology. Estimating the vulnerability of a building to hurricane damage—in order to determine the

[3] See https://www.sapling.com/7958883/average-cost-hurricane-insurance

monetary value of the damage—requires knowing how it was built, and how the building-materials will withstand the wind and water effects of the storms.

The scientific challenge of predicting hurricanes raises some surprising philosophical challenges. In 2016, we were approached by a team of scientists working for an insurance company who had been reading our previous philosophical work on scientific modelling and decision-making in the face of severe uncertainty. They asked for our help, and thus began a research collaboration on the philosophical challenges of insuring against hurricanes—the first ever, we would bet! In this article, we will share some of what makes hurricane insurance so philosophical interesting. We will look at how such insurance is done today, and why insurers were so unsatisfied that they brought in the philosophers.

4. How do you price a hurricane like Maria?

In most kinds of insurance (health, fire, theft, and so on), insurers calculate the likelihood of the event being insured against by looking at historical statistics. When buying an insurance policy, facts about you (or your house, car, etc.) are used to estimate how likely you are to experience the "event" you are insuring against (fire, theft, etc.). Your car insurance premium is calculated using things like your postcode, age, gender, and even the colour of your vehicle. Insurers looks at the statistics for burglary, accidents, and so on, for people in your area, of your age and gender, or with your colour car. From the insurer's point of view, they do this to try make it more likely that they'll make a profit from your insurance. From your point of view, it is important that they do this so that they don't run out of money before they can pay your insurance claim when you make it.

In the case of hurricanes, insurers can't do this, simply because there isn't enough data on destructive hurricanes: even in the USA, which has sophisticated records and experiences hurricanes in most years, there is too little data for actuarial modelling.[4] HURDAT2, the official database for hurricanes striking the Atlantic coast of the USA has ~300 storms to date and only 1/3 of those qualify as "major hurricanes". If the dataset is split by region, the numbers drop precipitously.[5] Compare this to the 6 million car accidents per year in the USA, and you can see why insurers have a much harder time with hurricanes than cars.

But if you live in a hurricane-prone region like Dominica, or south Florida, you *need* hurricane insurance. Insurers also know that, as destructive as Hurricane Maria was, such events are rare. They *should* be able to offer sustainable hurricane insurance, if only they can reliably fill the gap that the missing statistics usually play in their pricing process.

The probability of a hurricane hitting south Florida can be calculated in another way than the statistical approach used for car insurance: using scientific models. These models contain numerical representations of hurricanes: equations from physics and statistics that describe how the storms form, how they grow, and how they move across the Atlantic. In part, these models are based on what we know about the physics of hurricanes. For example, we know that tropical storms get the energy that makes them so ferocious from the sea: the warmer the sea-surface, the more energy is available to "drive" a large storm. But the models also use the

[4] For a discussion of hurricane modelling for insurance, see Shome *at al.*'s (2018), and for a discussion of hurricane risk the contributions to Collins and Walsh's (2019).

[5] We're referring to the number of datapoints in HURDAT2 for hurricanes that make landfall on the USA's Atlantic coast. The full database is at http://www.aoml.noaa.gov/hrd/hurdat/All_U.S._Hurricanes.html

statistics we have on where exactly in the Atlantic past storms formed, on periods of higher and lower hurricane activity, and so on.

The insurers treat the models as experts: they take the outputs of their models as an input for insurance-pricing, an input that they can't easily check because of the specialist knowledge required to produce it. They have to do this, if they want to avoid hiring their own scientists and building their own models, but it raises some problems.

Insurers know that scientists disagree; not only about some key questions of hurricane science, but also about how to take that science and put it into a computer simulation. The result is that there isn't just one model of hurricane formation in the Atlantic; there are many. When the Florida Commission on Hurricane Loss Projection Methodology, the industry regulator who licenses modelling firms, carried out their 2007 assessment of the modelling industry, they gathered a collection of 972 models![6,7]

Just knowing about these disagreements makes life hard for insurers, as they have no good way of choosing which model to buy. (Or which modelling company to hire.) The reasons for the disagreements are Greek to them, but any decision they make will boil down to a decision between the different models on offer. But insurers don't like the thought that they're implicitly making a choice on scientific or modelling questions they don't understand.

But if you are selling hurricane insurance, there is another option: you can hire a company like Risk Management Solutions (RMS), a leading modelling firm that uses a collection of thirteen models.[8] RMS's models represent (some of) the different views present in the scientific community. Rather than taking a stand on these disagreements themselves, they try to have a model for each major position. As an insurer, this is an attractive option: it offers you a way to stay out of scientific disputes. But, as we will now show, it doesn't quite solve the problem in the way insurers might hope.

5. Growing dissatisfaction with hurricane pricing

RMS uses multiple models because they understand the insurers' discomfort with scientific uncertainty. They also know that insurers need a definitive answer on how to price their products. And so, like many others in their position, RMS combines the outputs from their thirteen models into a just one number. So, any time an insurer asks the RMS software for the probability of a hurricane hitting a particular place, they don't see thirteen different probabilities, they see one: RMS's recommended view, called the Medium-Term Rate.

[6] The report is FCHLPM, "Report to the Florida House of Representatives Comparison of Hurricane Loss Projection Models," 2007, https://www.sbafla.com/method/Portals/Methodology/Meetings/2007/20071105_RubioReport.pdf. For a discussion of the 972 models, see Jayanta Guin, "Understanding Uncertainty," AIR Worldwide blog, 2010, http://www.air-worldwide.com/Publications/AIR-Currents/2010/Understanding-Uncertainty/.

[7] For discussions of model ensembles and their relation to uncertainty see Knutti's (2010), Marinacci's (2015), Parker's (2010, 2011, 2013), Stainforth, Allen *et al.*'s (2007a) and Stainforth, Downing *et al.*'s (2007b). For a discussion of probabilities in relation to ensembles see Frigg *et al.*'s (2015a) and Smith *et al.*'s (2014).

[8] Tom Sabbatelli and Jeff Waters, "We're Still All Wondering – Where Have All The Hurricanes Gone?," *The RMS Blog* (blog), October 27, 2015, http://www.rms.com/blog/2015/10/27/were-still-all-wondering-where-have-all-the-hurricanes-gone/.

The Medium-Term Rate is constructed by averaging the answers provided by each of RMS's thirteen models, in what is called a weighted average.[9] In a simple average, the answers are added together the answers and divided by 13. This gives each model equal say, or "weight", in the final answer. In a weighted average, some models can count more than others. RMS decides how much of a say to give to each model by scoring each one on how well it accounts for the past hurricane record.

They measure this by setting each model a "predictive test". The test asks them to calculate the hurricane activity during some historical period, like 1970-1975. When making these calculations, the models are only allowed to use data from before 1970. Each model's outputs for 1970-1975 is compared to the actual data for that period and given a score. The models are tested against many periods, and their overall scores are taken as a sign of each model's skill at predicting hurricanes. These scores then become weights in the average: the better a score, the more that output counts towards the overall answer.

The insurers aren't satisfied with this approach (Philp et al. 2019). That's partly because it is their job to worry about uncertainty, and to constantly strive to gain a better understanding of what they're insuring. But there are specific worries in the case of hurricane insurance, and model averaging in particular, that brought them to us.

The biggest problem is that averaging conceals important information from decision makers about just how uncertain the underlying science is. The averaging process, and the neat software packages that present the averaged results to insurers, focus attention on just one number. To the underwriter using the software, the underlying messy science, and the even messier reality, are swept under the carpet and they tend not to think about the fact that scientists disagree about important questions in hurricane science and modelling.

Although they can't understand the content of these disagreements, the people pricing insurance currently receive no information about them *at all*. But they could, we think, understand and make use of *some* facts about the disagreement. How "spread out" the results from different models are is important information, telling us something about the state of scientific knowledge about a question. The more spread out the results are, the more uncertainty and imprecision there is. This, we say, is valuable information that the decision-maker should use—and, importantly, can *use* even if the details of the disagreement are beyond their grasp. Later we will show how they can use it.

There are also two more technical problems with averaging model results. The first is that the predictive tests that are used to weigh each model's skill use historical data. This data, remember, is too little to use directly for insurance pricing—that is why we needed the simulation models in the first place. But here, the historical data is playing a key role, behind-the-scenes, in determining how the models are evaluated. It remains a weak link.

The second is that there isn't just one way to score the test. For predictive tests like these, there are many different "scoring rules". To see why, let's look at a simpler scenario. You want to know whether it will rain tomorrow, and so you check three different weather services. They say it is 30%, 50% and 80% likely to rain, respectively. Now if it does rain, how do we score those predictions? Some things seem straightforward: the highest prediction did best, because it did rain. But how much better did it do than the 50% prediction? On the one hand, 80% is 30

[9] Tom Sabbatelli, "Catastrophe Modeling - Part 2," *The RMS Blog* (blog), September 2, 2017, http://www.rms.com/blog/tag/catastrophe-modeling/page/2/; InsuranceERM, "RMS Responds to AIR's Attack on Hurricane Risk Modelling," Insurance ERM, May 29, 2018, https://www.insuranceerm.com/news-comment/rms-responds-to-airs-attack-on-hurricane-risk-modelling.html.

percentage points above 50%, which should play a role in measuring the difference between them. On the other hand, 50% is barely a prediction at all—it is what you might say if you didn't know anything at all. Shouldn't there be a bonus for "sticking your neck out"?

Statisticians disagree about how to answer these questions, with the result being that there are a great many competing rules.[10] These rules can disagree widely: ranking the predictions in completely different ways, and therefore leading to very different average answers.[11] This puts the insurer in a difficult position: they went to the modelling company because hurricane scientists disagreed, and they couldn't adjudicate that disagreement for themselves. But now it turns out the modellers themselves face a disagreement amongst still other experts: statisticians disagreeing over scoring rules. If insurers don't want their prices to reflect just one view of the science of hurricanes, they can reasonably say they don't want it to reflect just one view of the statistics of scoring predictions, either. But what can they do?

The insurers we work with worry about these problems. They try to compensate for them, "factoring in" their dissatisfaction by, for example, inflating the average probabilities. But while the worry is reasonable, they have no good way of choosing how much to inflate them by.

6. Less precision, more flexibility

Working with our insurance partners, we have developed a different way to price hurricane insurance.[12] It avoids these problems with averaging and gives insurance decision-makers a more flexible procedure for navigating scientific uncertainty.

Our approach starts with a simple thought: instead of trying to compress the disagreement between the three weather predictions (30%, 50%, 80%) down to one prediction (like the average, 53%), why not simply give a range? "It is 30-80% likely to rain tomorrow."

That's one way of dealing with the disagreement, but it creates its own questions.[13] The simplest question is: how exactly do we form the range? 30-80% includes all the predictions, but often if we consider *every* viewpoint we end up with unhelpfully wide ranges: if the probability is 30% there is no need for an umbrella, but if it is 80% one is practically required. If we consider all the options, we might end up *always* carrying an umbrella, which would be a nuisance.

This raises our second question: how do you make a decision when given a range like 30-80%? If you're told just one number, say 50%, it is simple. You think about your options (take an umbrella, don't) and consider what's likely to happen. If you take the umbrella, there's a 50% chance you carry it for no reason, and a 50% chance it rains and you use the umbrella to stay dry. If you don't take an umbrella, there's a 50% chance you enjoy your day unencumbered, and a 50% chance you get wet in the rain. The classic advice from economists is to "maximise expected utility"; in other words, to pick the option that does best on average. Some numbers can help to see how this works: let's say you really hate getting wet and

[10] See for example the list of rules at Australian Bureau of Meteorology, "Forecast Verification," 2017, https://web.archive.org/web/20171125111801/https://www.cawcr.gov.au/projects/verification/..

[11] For a discussion of this problem in the case of climate models, see Stainforth *et al.*'s (2007a).

[12] Our approach here reworks a recently developed decision theory called the *confidence approach* to tackle inputs from model ensembles. For a discussion of this approach see Bradley's (2017) and Hill's (2013, 2019).

[13] For a discussion of such "imprecise probabilities", the questions they raise, and how to make decisions with them, see S.Bradley's (2019).

represent that with the number -10. You don't like carrying an umbrella unnecessarily, but that's only -2. Staying dry using your umbrella is better than that, let's say +4. Finally, enjoying a sunny day unencumbered is best, +6. If you don't carry an umbrella, then you might enjoy a sunny day (50% x +6 = 3 unit of expected utility), but you might also get wet (50% x -10 = -5). So, your total expected utility is -2 if you don't take the umbrella. If do you carry the umbrella, there's a chance you do so unnecessarily (50% x -2 = -1) and a chance it keeps you dry (50% x +4 = 2), which gives a total of +1. You choose the act with the higher expected utility, so you take the umbrella.

But, if you have a range of probabilities for rain (30-80%) and a range for no rain (20-70%), you can't follow this advice. Economists and philosophers have offered various alternative rules for deciding with ranges of probabilities; one example is to be cautious and choose the option that does best in the worst-case scenario.[14] A problem for all of these rules is that they're one-size-fits all, and you might not want to commit to just one way of making your decision. The economist's rule about maximising expected utility is meant to be universal, and so are the alternatives for ranges. But in practice, people decide differently based on what is at stake. If we're deciding something trivial, like taking an umbrella when we leave the house, we might be happy to pick the middle of the range and decide using it. But if it is a life-and-death decision, we can't afford to ignore worst-case scenarios, and so we'd want to be more cautious and think about the full range.

7. The new method

Our method for making this kind of decision is designed to avoid these problems.

Here is a colourful example to show how it works. Suppose you are deciding whether to place a bet on your favourite contestant, let's call him Kevin, winning a dance contest. To place the bet you pay £50 upfront; if he wins you are paid back your £50 and receive another £50, if he doesn't win you lose your £50. So, you should expect this bet to make you money if the probability of Kevin winning is more than 50%.

But since you don't know the probability of Kevin winning, your situation is a bit like the weather predictions case. So, we will start off by working through various probabilities for him winning, and see which you accept. We start with the widest range: the probability of Kevin winning is between 0 and 100%. You should obviously believe that, but it is no help. We then narrow down the range, in various ways and see which you accept. This is based just on your subjective estimate of his chances, using your extensive experience in dance venues, your knowledge of how judges rule on performances, and your familiarity with Kevin's skills. Let's say that as we suggest various ranges to you, you accept the claim that his chance of winning is between 20-80%, and 30-50%, and so on, down to your best guess for a single number: 42%.

Now if we forced you to do this, it would be natural for you to protest that you aren't very sure at all about the precise number 42%. You're right. The crucial thing to realise is that your protestations hold the key to solving the problem. In fact, you will likely have protested about the 30-50% interval. In fact, every move toward a narrower interval may, and likely will, go hand in hand with decreasing confidence. The method asks you to pay attention to how your unease increases and keep track of your confidence in each of these claims, where "confidence" means how certain you feel that the chance of Kevin winning is in a certain range (or, in the case of 42%, is equal to that number). You should be less confident about the more specific

[14] For a review of various alternatives, see Heal and Millner's (2018).

claims; indeed, it would be incoherent to be more confident that the "right" probability is 42% than that it is between 30-50% because that range *includes* 42%!

You now have a ranking of your different estimates: most confident 0-100%, next most 30-80%, and so on down to least confident 42%. That tells us how the different ranges compare to one another. But, remembering our example of the umbrella decision and the life-threat decision, we want to get a sense of how confident you are in absolute terms. If your bet was for your life, rather than £50, you would want to be very confident in your probability estimate. Not just *more confident* than in some other guesses, but *confident enough to bet your life on*.

We can do this by putting the different ranges of probability you considered into groups, which we'll call Low, Medium and High confidence. Which category each range ends up in depends on how much *evidence* you used when you decided that the chance of Kevin winning was in that range. The more evidence, the higher the category. Imagine two people making this same bet. Joe is a dance aficionado; he goes to dance shows regularly and knows a lot about these competitions. He knows about each of the competitors, and how they've performed against each other before. Using that experience, Joe judges Kevin's chances to be between 30-50%. Roman, on the other hand, is new to dance. He hasn't been to a dance show before and has only just heard of Kevin. He also judges Kevin's chances to be between 30-50%, but it is an uninformed guess. Joe's judgement would get High confidence, but Roman's only Low.

Let's say that your experience with dance competitions, and what you know about Kevin, is enough to justify categorising your 30-50% estimate as Medium confidence; your 20-80% estimate as high confidence; and your 42% estimate as low confidence.

Why do all this? To come to a decision two further issues are crucial: the importance the decision has to you and your attitude to risk. The decision of taking an umbrella is trivial and hence not very important, but if you're shot if you get it wrong, then the decision really matters. How you decide will also depend on whether you value safety or whether you are willing to take risks. An adequate decision algorithm must take these two factors into account, and having various ranges of probabilities ranked by the confidence you attach to them helps you to do so. We now see how.

The "stakes" of a decision reflect how important the decision is to you. We're going to think of stakes as a number on a 0 to 1 scale, where 0 is totally unimportant and 1 is the most important decision you can imagine. There are different ways we could measure the stakes of a decision, but for now we'll simplify and say that for you the importance of a decision is determined by the worst possible outcome. So in the example of the dance competition the "stakes" are determined by the money you stand to lose: the £50. How important is losing £50 to you? If you are poor, it might be very important! If you're rich, you might not even notice it. For the moment let us say you think this is a moderately important decision and assign it stakes $s = 0.5$.

Second, we want to know how cautious you are. If Joe is very cautious, he might want to be very confident when he makes choices, even if there isn't much at stake. Roman, who is much bolder, is willing to make even important decisions using little evidence. What we want to know for your bet is your answer to the question "how much confidence do you need in order to make moderately important decisions, with stakes around 0.5?" The answer to the question should be one of our confidence levels: Low, Medium, or High. Joe, who is very cautious, will answer High. Roman, who is gung-ho, will say Low.

Let's say that you fall in between them again, and you answer: Medium. This tells you which one of your probability estimates you should use, based on your caution and the stakes (the importance of the decision to you). When we categorised your probability estimates, we put 30-50% into the Medium category, so you should use 30-50% to make your decision.

This is our answer to the first problem for deciding using ranges: which range? The answer is: the range that fits your desire for confidence, which is based on how cautious you are and what is at stake.

You can now use one of the rules that economists and philosophers have suggested for deciding using a range. Let's keep things simple and say you'll use the cautious rule, by choosing the options that you expect to do best if things turn out for the worst, from your perspective. This rule is called "maximin" expected utility. Here's how it works: you start with one of the options (e.g., bet on Kevin), and calculate the expected utility of that option using each of the probabilities in the range (which here is 30-50%). You're looking for the *minimum* expected utility the option has. So, you start by working out the expected utility for 30% (for the moment we'll pretend money and utility are the same): 30% x 50 + 70% x -50 = -20. You do the same for all the probabilities in the range, working up to 50% where the expected utility is 0. The minimum of all these values is -20, so that's the only number you need to care about for the "bet" option. You then do the same for the "don't bet" option. This is much simpler as you don't lose or gain anything, so it is always 0. You then compare these minimum values and choose the option with the *highest* worst-case utility (hence, maximin). In this case, that means "don't bet" as 0 is greater than -20.

You would only expect to make money betting on Kevin if the probability of her winning is over 50%. As it gets lower than that, you expect to lose more and more, so because you think the probability of her winning is in the range 30-50%, it is a bad bet! You don't expect to make money based on what you think the probabilities are, and there's only one way you could just break even: if things turned out for the *best,* and the probability was at the very top of the range you think it is in. So, you don't bet.

8. Insurance pricing

That all seems like a lot of effort for betting on a dancer. But in our more complicated insurance context, the machinery of confidence, caution and stakes does important and useful work.

Let's go back to talking about hurricane insurance. Let's imagine an insurer who wants to sell a single insurance contract on house damage due to hurricanes. This is their first contract, and it is for insurance against an event, which we'll label E: "a hurricane strikes Fort Lauderdale in 2025". The contract is for a total value of £100,000 and it is a binary contract: it pays out either £0 if the event does not occur, or £100,000 if it does (there are no intermediate values for partial damage).

Pricing insurance involves working out how much you need to charge to make a profit, in a way that is not too different from deciding whether to place a bet. As in any business, you make a profit in insurance when you take more money in than you pay out. So, the annual price charged for a contract (money in) needs to be larger than the sum of two basic expenses: the expected pay out to customers claiming on their insurance contracts, and the cost of holding money so that it is available to pay those customers if the need arises. Both of these depend on the probability of the hurricane striking Fort Lauderdale. For payments out, that's obvious: you pay out when the hurricane hits. For money held, that's less obvious but it is because the insurance regulator requires insurers to hold capital according to a formula which uses the probability that they will pay out.

Here is an example. If the probability of the hurricane is 1%, then you should expect to pay 1% of the £100,000 insured in any year. That's the first expense: an expected pay-out of £1,000. For the cost of capital holdings, we'll simplify and say that the insurer needs to hold the whole £100,000. (This sounds reasonable: after all, if I claim on the contract, I want them to have all £100,000! But in reality, insurers have many customers and they only ever hold a fraction of the money they would need to pay out if everyone claimed at once.) If the cost of capital is 5%, then the second expense, cost of holding capital, is 5% of £100,000 or £5,000. So, to make a profit, the insurer must charge more than £6,000 per year.

That tells us how insurance is priced if we know the probability of the event. But our original problem was: it is very hard to figure out these probabilities!

Our insurer will estimate this probability by talking to some scientific modellers. Let's imagine that, like Risk Management Solutions, they use a collection of models to advise the insurer. Table 1 shows how the information that the insurer gets looks like: it's a list with 13 probabilities, each provided by one model constructed by the scientists. The list also contains each model's weight, which is, as we have seen, obtained by scoring each model in a predictive test. In our example Model 1 scored best and gets a weight of 23.7%. Model 10 scored worst and is only going to count for 1.6%.

Table 1. "Model outputs" for toy example

Model	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
$p(E)$	.0070	.0071	.0068	.0074	.0076	.0061	.0083	.0086	.0091	.0092
Weight (%)	23.7	20.7	15.8	11.6	11.5	7.3	3.2	3.0	1.7	1.6

Remember that in the "standard" approach, insurers will calculate a weighted average of these probabilities. Using those weights, the weighted average probability is 0.0072. We can now price this contract, just as we did above if we take this average to be the probability for the event of a hurricane striking Fort Lauderdale to occur. The first expense, expected pay outs, is 0.0072×100,000=£720. The second expense will be the same as before, £5,000. So the minimum price is £5,720. As we noted above, concerned insurers often inflate these averages for "safety". This might be as crude as doubling the probability from 0.0072 to 0.0144. Going through the calculations with that probability, we get £6,440.

9. Pricing with Confidence

Pure averaging and averaging with these ad hoc adjustments are unsatisfactory methods. The confidence approach offers an attractive alternative. To apply the confidence approach, we start with forming all the ranges of probability that the decision-maker accepts. The insurer does what you did in our dance example: it gathers all the evidence it can get. This includes all the model results, and what the skill score says about them. But it is not restricted to that: the insurer can take different skill scores into account and see whether they agree; or consider the nature of the different models in the ensemble and how they have been constructed; or weigh up the nature of the scientific disagreements (are they disagreements over principles, or over the application of principles, or over parameter values, or over the use of numerical techniques, or …); or it can talk to different scientists, including those involved in the construction of the different models, to get the full picture of the state of play in the field. Based on all the evidence gathered, the insurer will then construct various intervals and form an opinion about how confident it can be in these intervals.[15]

We hope this sounds like common-sense to you. But notice that common-sense is not how things are currently done. Alternative skill scores are not considered when calculating weighted averages, and no other sources of information about either the models, the modelling process, or the state of knowledge in the field is taken into account. Opening up to these issues will allow insurance scientists to form a much more nuanced picture of the available evidence for and against various scenarios, and come to better founded judgment about the trustworthiness of model outputs, than they could by uncritically relying on mechanically calculated skill scores.

Assume now that the science team of the insurance has gone through this process and, considering all available evidence, has come to the conclusion that it should consider three intervals: [0.007,0.0071] with confidence level Low, [0.0068,0.0076] with confidence level Medium, and [0.0061,0.0091] with confidence level High.

The insurer now has to think about the stakes involved in this decision. This contract is the insurer's first; it will be their whole business and so the risk of going bust is high. Still, no one's life is at stake and there is no impact on anything else (e.g., no other business which might be taken down). So, the insurer concludes that their stakes are moderately high, $s=0.75$.

Next in line is cautiousness. As insurance against natural catastrophes involves significant uncertainty, this insurer can't be too cautious. So, let us suppose that they only demand High confidence for important decisions of stakes 0.9 or higher. For a decision like ours, with stakes 0.75, they're happy with Medium confidence.

Now we look at the intervals above and see that the relevant interval for that confidence level is [0.0068, 0.0076]. The insurer can now apply the same cautious decision rule we used before, maximin expected utility, and make the choice that turns out best if things go as badly as possible. In insurance, *higher* probabilities are worse, because they make it more likely you need to pay out. So, they care only about the largest probability in this range, 0.0076. The first expense, expected pay out, is 0.0076×100,000=£760. The holdings are exactly as before. So we end up with a minimum price of £5,760.

Pricing method	Price
Averaging	5,720
Averaging + "safety" factor	6,440
Confidence	5,760

It is now interesting to compare the three methods. The "pure" average price is £5,720, but as we saw earlier, it has a lot of problems associated with it. The "safety" price, which involves ad hoc adjustments to the average, is much higher, £6,440. Our new confidence price is only a

[15] The details of this are discussed in our (2021).

little higher than the average at £5,760. But to get there, the insurer had to follow a completely different procedure. They used the full set of model outputs. The price depends on how important the decision is, and on how cautious the insurer is. In other decisions, they will get completely different answers. On the other hand, the average and safety prices are always calculated the same way.

10. Better decision-making

Insurance is meant to put a "price on risk" so that people can pay protect themselves from the unexpected. Insurers would also like to put a price the kind of uncertainty that we discuss here: not knowing what the probability of some event is. The current options available to insurers are all ad hoc and there is no guarantee for insurance companies that their staff are responding to different risks (hurricane, earthquake, wildfire) in a common and systematic way.

Our approach allows insurers to systematically set this kind of "uncertainty premium". For an insurance company, three of the ingredients discussed above will be a matter of policy. They will need to agree on a way of measuring stakes that allows them to compare the different kinds of decisions they make and decide which are more and less important. Cautiousness can similarly be determined by a high-level decision about how much confidence to demand for decisions of various stakes. Finally, a decision rule will need to be selected; either maximin expected utility or one of its competitors.

With these three elements in place the cautious confidence approach provides a recipe for pricing insurance that is sensitive to all of the evidence available for that decision and which responds naturally (through the cautiousness and stakes) to the different nature of each decision taken. This kind of flexible but systematic treatment of uncertainty is what insurers tell us they have been missing in catastrophe insurance.

The last two paragraphs also demonstrate a second benefit: the confidence approach fits naturally with the kind of distributed decision-making and corporate responsibility found in large insurance companies. The different parts of the recipe are naturally provided by different stakeholders. The cautiousness function is ultimately determined by the shareholders' appetite for uncertainty. The way of determining stakes will be set by senior management in charge of portfolio management and capital allocation. The probability functions themselves, along with their nesting and grouping into confidence levels, come from the science department who cover that particular risk. This is a better fit with how insurance decision-making should work than having underwriters "adjust" scientific estimates of probability individually.

Finally, looking beyond the insurance case, the method we outlined is applicable in all instances of decision-making under uncertainty. This is because nothing depends on the evidence being provided in form of an ensemble of models. The multiple sources of information be experts who hold diverging views, or they can be a mixture of different kinds of sources.[16] The "distributed" nature of the implementation can be an advantage also in other contexts, where different stakeholders will the probability functions, and determine the stakes, and set the cautiousness function, which can be a beneficial setup in many social situations.

[16] For a discussion of expert elicitation see Cooke's (1991), Morgan's (2014) and the contributions to Martin and Bouman's (2014). For a discussion expert elicitation in the context of climate change adaptation see Thompson et al.'s (2016).

Bibliography

Bradley, R. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.

Bradley, R. and K. Steele 2015. Making Climate Decisions. *Philosophy Compass* 10: 799-810.

Bradley, S. 2019. Imprecise Probabilities. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities/.

Collins, J. M. and K. Walsh (eds.) 2019. *Hurricane Risk*. Cham: Springer.

Cooke, R. M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford: Oxford University Press.

Frigg, R., L. A. Smith, and D. A. Stainforth 2015a. An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of Ukcp09. *Synthese* 192: 3979-4008.

Frigg, R., E. Thompson, and C. Werndl 2015b. Philosophy of Climate Science Part I: Observing Climate Change. *Philosophy Compass* 10: 953-964,.

Frigg, R., E. Thompson, and C. Werndl 2015c. Philosophy of Climate Science Part II: Modelling Climate Change. *Philosophy Compass* 10: 965-977.

Heal, G. and A. Millner 2018. Uncertainty and Decision-Making in Environmental Economics: Conceptual Issues. In P. Dasgputa, S. Pattanayak, and V. K. Smith (eds.), *Handbook of Environmental Economics Vol 4*, pp. 439-468.

Hill, B. 2013. Confidence and Decision. *Games and Economic Behavior* 82: 675-692.

Hill, B. 2019. Confidence in Beliefs and Rational Decision Making. *Economics and Philosophy* 35: 223-258.

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

Knutti, R. 2010. The End of Model Democracy? *Climate Change* 102: 395-404.

Marinacci, M. 2015. Model Uncertainty. *Journal of the European Economic Association* 13: 1022-1100.

Martini, C. and M. Boumans (eds.) 2014. *Experts and Consensus in Social Science*. NewYork: Springer.

Morgan, M. G. 2014. Use (and Abuse) of Expert Elicitation in Support of Decision Making for Public Policy. *PNAS* 111: 7176-7184.

Parker, W. S. 2010. Predicting Weather and Climate: Uncertainty, Ensembles and Probability. *Studies in History and Philosophy of Modern Physics* 41: 263-272.

Parker, W. S. 2011. When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science* 78: 579-600.

Parker, W. S. 2013. Ensemble Modeling, Uncertainty and Robust Predictions. *Wiley Interdisciplinary Reviews: Climate Change* 4: 213-223.

Parker, W. S. 2018. Climate Science. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/sum2018/entries/climate-science/.

Philp, T., T. Sabbatelli, C. Roberston, and P. Wilson 2019. Ssues of Importance to the (Re)Insurance Industry: A Timescale Perspective. In J. Collins, and K. Walsh (eds.), *Hurricane Risk*. Cham: Springer, pp. 1-22.

Roussos, J., R. Bradley, and R. Frigg 2021. Making Confident Decisions with Model Ensembles *Philosophy of Science* 88: 439-460.

Shome, N., M. Rahnama, S. Jewson, and P. Wilson 2018. Quantifying Model Uncertainty and Risk. In G. Michel (ed.), *Risk Modeling for Hazards and Disasters*. Amsterdam: Elsevier, pp. 3-46.

Smith, L. A., H. Du, E. B. Suckling, and F. Niehörster 2014. Probabilistic Skill in Ensemble Seasonal Forecasts. *Quarterly Journal of the Royal Meteorological Society* 140: DOI:10.1002/qj.2403.

Stainforth, D. A., M. R. Allen, E. R. Tredger, and L. A. Smith 2007a. Confidence, Uncertainty and Decision-Dupport Relevance in Climate Predictions. *Philosophical Transaction of the Royal Socity A* 365: 2145-2161.

Stainforth, D. A., T. E. Downing, R. Washington, A. Lopez, and M. New 2007b. Issues in the Interpretation of Climate Model Ensembles to Inform Decisions. *Philosophical Transactions of the Royal Society A* 365: 2163-2177.

Thompson, E., R. Frigg, and C. Helgeson 2016. Expert Judgment for Climate Change Adaptation. *Philosophy of Science* 83: 1110-1121.