

ARTICLE

# Confidence in Probabilistic Risk Assessment

Luca Zanetti 

Department of Humanities and Life Sciences (SUV), Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy

Email: [luca.zanetti@iusspavia.it](mailto:luca.zanetti@iusspavia.it)

(Received 29 April 2023; revised 14 October 2023; accepted 06 November 2023; first published online 15 November 2023)

## Abstract

Epistemic uncertainties are included in probabilistic risk assessment (PRA) as second-order probabilities that represent the degrees of belief of the scientists that a model is correct. In this article, I propose an alternative approach that incorporates the scientist's confidence in a probability set for a given quantity. First, I give some arguments against the use of precise probabilities to estimate scientific uncertainty in risk analysis. I then extend the “confidence approach” developed by Brian Hill and Richard Bradley to PRA. Finally, I claim that this approach represents model uncertainty better than the standard (Bayesian) model does.

## I Introduction

The assessment of natural hazards is subject to two types of model uncertainty. On one hand, there are *structural* uncertainties, which concern the form of the equations used to model natural phenomena (e.g., uncertainty whether all the relevant explanatory variables have been included in the model). On the other hand, there are *parametric* uncertainties, which concern the value of the parameters in these equations (e.g., uncertainty if the values of these parameters have been measured correctly). Both these uncertainties are typically due to a shortage of data: historical catalogs of rare and extreme natural events (such as strong earthquakes, volcanic eruptions, and tsunamis) often comprise only a few centuries of recordings, whereas a complete validation of natural hazard models would require thousands of years of data. This last point deserves some emphasis: given that the “true” data-generating processes are unknown, epistemic uncertainties are often *severe*, that is, they cannot be quantified probabilistically in any meaningful way.

The target of this article is the current practice in probabilistic risk assessment (PRA) to include these uncertainties as probabilities defined over an ensemble of models. Probability theory is used twice in PRA: first, to quantify uncertainties that are due to the intrinsic randomness of natural phenomena (first-order uncertainty), and second, to quantify the uncertainty about these first-order estimates (second-order uncertainty). These second-order probabilities often reflect the judgments a panel of experts gives that evaluate each model's reliability in a specific case. Given our remarks about severe uncertainty, these probabilities are interpreted as the

subjective degrees of belief of the experts that the model is correct rather than as the objective probability that the model is the correct one. The overall hazard is expressed by the frequency of exceedance of an adverse event. This frequency is calculated as the weighted average of the estimates that the models produce, with the weights corresponding to these models' probabilities.

I propose an alternative approach that incorporates the scientist's confidence in a probability set for a given quantity. My proposal extends the "confidence approach" formulated by Hill (2013, 2019) and developed by Bradley (2017) and Roussos, Bradley, and Frigg (2021). This approach is promising for three reasons. First, it distinguishes between the "weight of evidence" in favor of a forecast from the forecast itself. For example, new data about frequent but low-intensity events can make experts more confident that a particular model is correct, even though their best estimate of the frequency of these events remains the same. Second, confidence levels (e.g., high, medium, and low) are qualitative and are therefore distinct from degrees of belief. This reflects the current practice of eliciting the individual judgments of experts as qualitative assessments rather than as probabilities. Third and finally, the confidence in a probabilistic forecast required for that probability to play a role in a decision depends on what is at stake in that decision. This corresponds to how risk analysis, and especially PRA, is currently practiced. Indeed, the evaluation of the models may be influenced by practical goals and aims and does not depend merely on the scientific accuracy of the models themselves.

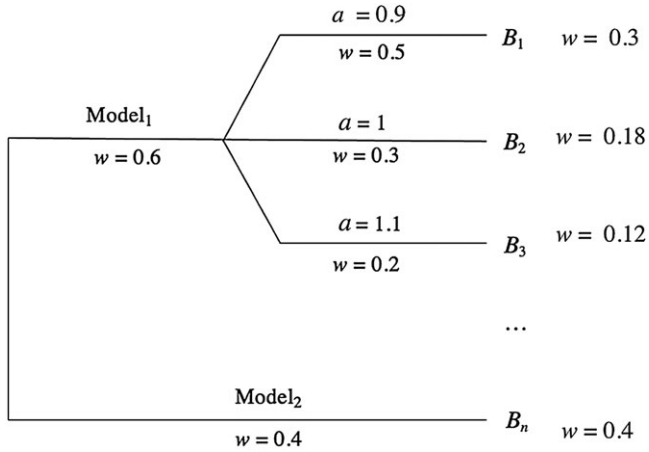
I proceed as follows. In section 2, I introduce the method of logic trees for estimating natural hazards. I then describe two alternative ways to include epistemic uncertainties in PRA: the standard (Bayesian) approach (section 3) and the confidence approach (section 4). The standard approach is *bottom up*: the degree of belief of the experts that an estimate is correct is calculated by multiplying the weights that the experts assign to the individual components of the models. The confidence approach is *top down*: experts are tasked with assessing their confidence directly about the frequency of the event of interest. In section 5, I show that these two approaches give different results in a concrete case (the New Italian Seismic Hazard Map [MPS19]). In section 6, I extend the confidence approach to PRA. I argue that model uncertainty should be represented using confidence rankings with multiple foci and show that this representation can be made consistent with both Hill's and Bradley's formulations. In section 7, I compare the confidence approach with the Bayesian approach. I claim that the confidence approach is the one that best represents model uncertainty. Section 8 concludes the article.

## 2 Logic trees for estimating natural hazards

Epistemic uncertainties are included in PRA using *logic trees* (see, e.g., Baker, Bradley, and Stafford 2021, sec. 6.7, 272–75).<sup>1</sup> Setting up a logic tree requires two steps. (1) The analyst selects all the models that can describe the area of interest, together with different values for their parameters. This selection is often based on a literature review, but recently, some studies have included models whose formulation was

---

<sup>1</sup> Epistemic uncertainties are due to limited data or an insufficient understanding of natural and social phenomena; by contrast, *aleatoric* uncertainties are due to the intrinsic variability of these phenomena.



**Figure 1.** A simplified logic tree.  $B_i$  is the  $i$ th branch of the logic tree,  $w$  is weight, and  $a$  is the model parameter.

solicited directly for the study (Meletti et al. 2021). (2) Each model is assigned a *weight*, expressed by a probability, that represents the analyst’s confidence that the model is correct (Scherbaum and Kuehn 2011; Bommer 2012; Musson 2012).

Logic trees can be understood as representations of a sequence of modeling choices (Kulkarni, Youngs, and Coppersmith 1984). The various stages in the construction of the models are reflected in the order of the nodes. For example, a particular model is selected, then all the values for its parameters are included. At each stage, the probability assigned to a branch departing from the node corresponds informally to the probability that the choice is correct conditional on the assumption that all previous choices were correct. Figure 1 is a simplified logic tree. The logic tree depicted in the figure includes two models. The weights assigned to these models are, respectively, 0.6 and 0.4. The first model has a parameter  $a$ ; the branches that depart from the upper node represent three different choices for the value of  $a$  and their weights.

In PRA, hazards are expressed as frequencies of exceedance,<sup>2</sup> that is, the annual probability of an event whose intensity exceeds a specific threshold at the site of interest (e.g., 10 percent in 50 years, corresponding to a return period of 475 years<sup>3</sup>). Each complete model, corresponding to a branch of the logic tree, calculates the probability of exceedance of each intensity level in the range of interest included between the minimum and maximum intensities that are considered for the study. The overall hazard at the site is calculated using the total probability law as the sum of all exceedance frequencies estimated by the models in the ensemble, each weighted by the probability of the model.

Finally, logic trees can include both structural and parametric uncertainties. Let’s give an example. It is usually assumed that earthquake magnitudes follow the

<sup>2</sup> In risk analysis, the hazard (the probability of an adverse event) is often distinguished from the *vulnerability* (the probability of a loss were the adverse event to occur) and the *exposure* (the potential loss).

<sup>3</sup> The return period of an event with a specific intensity is the converse of its frequency.

Gutenberg–Richter distribution, according to which the number of events with magnitude  $m$  in a specific area is  $10^{a-bm}$ , where  $a$  and  $b$  are constants (the “ $a$ -value” and the “ $b$ -value,” respectively) that characterize the site. The  $b$ -value, in particular, expresses the ratio between low-intensity and high-intensity seismic events. However, some studies show that this equation does not always hold in the proximity of faults. In this case, different equations have been used, for example, characteristic earthquake models.<sup>4</sup> If the analyst is uncertain about the geophysical features of the site, and in particular, whether there are active faults, the logic tree may include structurally different models, for example, a model that incorporates the Gutenberg–Richter equation (model 1 in figure 1) and a fault-based model (model 2). Among the models that do include the Gutenberg–Richter equation, there can be a difference in the estimate of the  $b$ -value at the site (e.g., 0.9, 1, or 1.1), and these different values can be included as branches of the logic tree.

### 3 Model ensembles and the use of experts in probabilistic risk assessment

The interpretation of the weights in the logic tree is Bayesian in two senses. First, the weights represent “subjective estimates for the degree-of-certainty or degree-of-belief—expressed within the framework of probability theory—that the corresponding model is the one that should be used” (Scherbaum and Kuehn 2011, 1238). In larger studies, these weights do not correspond to the judgments of a single analyst but reflect the options expressed by a panel of experts.

The use of experts in PRA has been influenced mostly by a report of the US Senior Seismic Hazard Analysis Committee (SSHAC) from 1997. This report distinguishes between individual proponents, which formulate the models; experts, who evaluate these models; and a “technical integrator” (TI) or “technical facilitator/integrator,” who oversees the overall study. The SSHAC report also emphasizes that the weights should be the product of a *consensus* among the experts on the panel (23, 25–26). The report distinguishes between four types of consensus (from type 1 to type 4). What is expected is that all experts agree that a particular composite probability distribution represents them as a group (type 3) or, more weakly, that the distribution represents the overall scientific community (type 4). It is not required, however, that each expert believe in the same probability distribution for a random variable or model parameter (type 2) or that each expert agree with all other experts that a given model or a given value for a parameter is correct (type 1).

The report describes a three-stage procedure to achieve the required level of consensus (Senior Seismic Hazard Analysis Committee [SSHAC] 1997, 23, table 3-1; see also Budnitz et al. 1998; Kammerer and Ake 2010). In the selection phase, the TI reviews the literature and selects the models that will be included in the study. In the evaluation phase, the TI forms a panel of experts that evaluates the models. Scientists can participate in the process in two roles, both as proponents (of their models) in the first phase and as evaluators (of others’ models) in the second phase. Finally, in the aggregation phase, the TI formulates the weights based on the judgments of the panel of experts and builds the logic tree. The report stresses that, even though the

<sup>4</sup> A characteristic earthquake model characterizes the seismicity of the site in terms of the periodic occurrence of a single event with a specific intensity.

individual scientists have scientific responsibility for their models, the TI has responsibility for the final result (SSHAC 1997, appendix, 16).

The second sense in which the standard approach is Bayesian is that the weights of the logic tree can be updated as new data are collected (Secanell et al. 2018). Given an observation  $A$  consisting of recordings of the activity at the site (e.g., seismic activity) in a specific interval of time (e.g., one year), the probability that the model that corresponds to the  $i$ -th branch of the logic tree is correct can be calculated, using Bayes's law, by setting the prior probability  $\Pr(B_i)$  to the weights elicited from the experts,  $\Pr(A|B_i)$ , to the frequency of  $A$  calculated by the model, whereas  $\Pr(A)$  is treated as a normalizing factor calculated by the total probability law as the sum of the probability of observing  $A$  if the model  $B_j$  is correct, weighted by the probability that  $B_j$  is correct. The posterior probability that a model is correct given some set of observations can be used to score the predictive power of the models with respect to an independent set of data and conditional on the judgments of experts (Selva and Sandri 2013). More in general, the posterior probability of each branch corresponds to the degree of confidence that the experts should have that the model is correct in light of the new data.<sup>5</sup>

The prior probabilities assigned to the models, on the other hand, have to be precise. The reference value in engineering is indeed the mean hazard, which is the easiest both to compute and to include in an analysis that compares the mean economic losses (McGuire, Cornell, and Toro 2005). As Baker, Bradley, and Stafford (2021, 276) recently noticed, “from a practical perspective, the mean hazard . . . is the nearly universal choice for summarizing the results” of PRA. Mean hazards are the sum of all the estimates  $v_i(a)$  multiplied by the weight  $w_i$  assigned to the branch  $B_i$  of the logic tree, and the weight of each branch of the logic tree, corresponding to a complete model, is calculated by multiplying the weights of its segments. Therefore the experts should be able to express their evaluation of each subcomponent of the different models by assigning a precise weight. Alternately, hazards are estimated using a percentile (usually, the 85th or the 90th percentile) of the estimates produced by the models in the ensemble (Abrahamson and Bommer 2005). Again, it is assumed that the weights assigned to the models can be expressed as precise probabilities (Musson 2005).

The use of precise probabilities to represent model uncertainty has, however, become the subject of controversy, especially in the climate sciences. In particular, the reports of the Intergovernmental Panel on Climate Change (IPCC) have gradually shifted from expressing uncertainty in probabilistic terms to using qualitative levels of confidence (Mastrandrea et al. 2011; Aven and Renn 2015; Wüthrich 2016). To mention just one notable example, a guidance note for the IPCC Sixth Assessment Report of 2021–23 (AR6) defined *risk* as a potential for adverse consequences (Reisinger et al. 2020). These consequences include, for example, life losses; negative economic and social impacts; economic losses; losses in infrastructure, ecosystems, and species, and so on, whereas the word “potential” indicates uncertainty about their realization. The note stresses that “the definition does not require the adverse

---

<sup>5</sup> Notice that this is still rarely done in practice. By contrast, some recent PSHA assigns a weight that combines the experts' judgments with a scoring of the forecasting performance of the models with respect to a set of independent data; see section 5.

consequence or the degree of uncertainty or likelihood of those consequences to be quantified” (8). On the contrary, AR6 adopts three different language scales to express uncertainty about the estimates of these losses: an evidence/agreement scale (e.g., *robust evidence, high agreement*), a confidence scale (ranging from *very low* to *very high*), and a likelihood scale (ranging from *exceptionally unlikely* to *virtually certain*). The likelihood scale is explicitly tied to intervals of probabilities (e.g., *very likely* corresponds to 90–100 percent probability). However, the guidance notes emphasize that confidence should not be interpreted probabilistically (Janzwood 2020, 1657–58). Climate projections are often similar to risk analysis insofar as historical data sets are limited. However, the distinction between different scales for representing uncertainty does not seem to have percolated into PRA. By contrast, model uncertainties are still expressed in PRA exclusively as second-order probabilities.

The use of model weights can be defended in three ways. The first response is that even if historical data are scarce, and so there is large epistemic uncertainty about which model is correct, the weights can be precise because these weights correspond to the experts’ estimates of their uncertainty. However, this response is unsatisfactory, because the problem is exactly whether the available evidence allows precise credences. The weights assigned by the experts may indeed be sensitive to available evidence (Klügel 2011, 40). But as Bradley, Helgeson, and Hill (2017, 509) emphasized, for example, “if [a person] has trouble forming precise first-order probabilities, why would he have any less trouble forming precise second-order confidence weights?”

The second response is that the weights are a precise mathematical representation of the experts’ credences, which can themselves be imprecise. Indeed, the experts on the panel often provide only a qualitative assessment of the likelihood of the models (Kammerer and Ake 2010). The precise weights are then inferred from these judgments, and aggregators typically have some choice of how to represent them in the form of a probability distribution (Marzocchi and Jordan 2017). This response is standard in the literature on structured expert judgment elicitation (e.g., O’Hagan et al. 2006). However, if the experts do not express probabilistic judgments, it is arbitrary, to some extent, which probability corresponds to which level of confidence. However, given that the overall hazard is calculated as the weighted average of the estimates produced by the models in the ensemble, these choices can lead to either overestimating or underestimating the hazard.

The third, and more promising, response concedes that the experts’ judgments cannot indeed be precise while insisting, at the same time, that if epistemic uncertainty cannot be reduced, then it must be communicated as part of the results of the risk analysis. The first fact imposes limitations on how this uncertainty can be represented. In particular, epistemic uncertainty should not be represented by precise second-order weights but can be represented using *imprecise probabilities*. In what follows, I present one such approach that incorporates the analyst’s confidence that the “true” frequency lies within a given probability interval. My approach extends Hill’s and Bradley’s “confidence approach.” This approach has already been applied to clarify the relationship between qualitative and quantitative metrics in the IPCC reports (Bradley, Helgeson, and Hill 2017; Helgeson, Bradley, and Hill 2018). However, it has not yet been applied to PRA. My aim is to answer two questions: (1) Can the confidence approach be applied to PRA? and (2) If so, does this approach provide a better representation of model uncertainty than the standard one?

#### 4 Credences and confidence

In this section, I introduce the formal notion of confidence.<sup>6</sup> My presentation follows Hill (2013, 2019) and Bradley (2017). Confidence, as understood here, is a propositional attitude, that is, an attitude that an agent can have toward a proposition. The propositions that will be considered for our purposes are forecasts, which usually take the form of probability judgments, that is, judgments that attribute a probability to an event. The agent's confidence in a probability judgment corresponds to "one's attitude of being more or less sure of one's belief" (Hill 2019, 4).

Confidence must be distinguished from the forecasts themselves. The latter estimate the objective probability or chance of an event (e.g., the probability of an event whose intensity exceeds a specific threshold), whereas the former captures the experts' individual or collective attitudes. Confidence should also be distinguished from second-order probabilities. The latter are representations of confidence as a numerical degree of belief. If such a representation is not warranted, as argued earlier, then one can either renounce the notion of confidence altogether or consider an ordinal notion, that is, a confidence relation that induces an order on the set of probability judgments without assigning a second-order probability to these judgments.

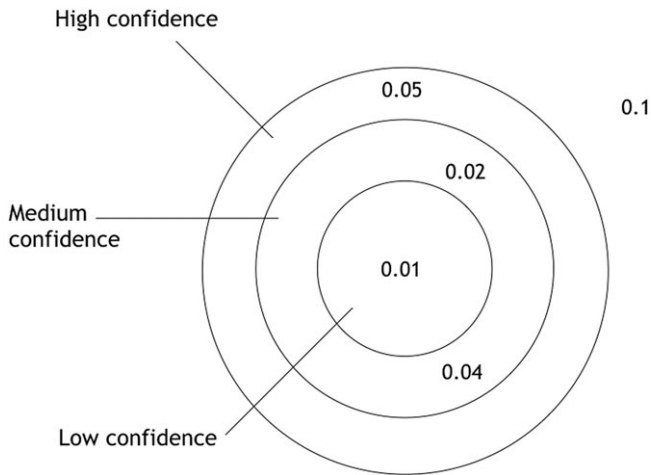
Hill (2013, 2019) formulates an ordinal notion of confidence. Hill proposes that the agent's confidence can be represented by a nested family of sets of probability measures. A probability measure assigns a probability to each member of a set of events. Sets of probability measures are often used to represent imprecise credences, that is, cases in which the agent's belief is better represented by a probability interval than by a precise probability. A nested family of sets is a set that contains a chain of subsets. Figure 2 is a representation of Hill's approach. In figure 2, the judgment that  $\Pr(A) \leq 0.05$  holds in all the sets in the ranking. Therefore the expert is highly confident that  $\Pr(A)$  is 0.05 or lower. At the same time,  $\Pr(A) = 0.1$  is outside the confidence rank of the expert, which means that the expert has no confidence that that is the correct estimate of  $\Pr(A)$ .

Three features of Hill's proposal are worth mentioning (Hill 2013, 678; 2019, 229–31). First, larger sets in the ranking (i.e., outer sets) correspond to higher levels of confidence, whereas smaller (i.e., inner) sets correspond to lower levels of confidence. Each set in the ranking represents all the values of  $\Pr(A)$  that the agent holds with the same level of confidence. For example, the agent can have high confidence that the "true" value lies within a given set  $\mathcal{A}$  but only medium confidence that it lies within a smaller set  $\mathcal{B} \subset \mathcal{A}$  and even less confidence that  $\Pr(A)$  lies in an even smaller set  $\mathcal{C} \subset \mathcal{B}$ .

Second, each set in the confidence ranking can be "lifted" to a set of probability judgments. The set of probability judgments that correspond to a given set in the confidence ranking is the set of judgments that hold for all probability measures in that set. The probability judgments corresponding to a given rank are the judgments

---

<sup>6</sup> This notion is closely related to the use of confidence intervals in classical statistics, as emphasized, for example, by Hill (2013, 676): "Confidence intervals at a standard fixed level (for example, the 95% level) are used to inform decisions. The proposed decision rule would suggest that one may vary the confidence intervals one uses in decisions: whereas if relatively little is at stake one may use confidence intervals at, say, the 90% level, if there is a lot at stake one might insist on relying only on intervals that are at, say, the 99% level."



**Figure 2.** Hill's (2013) representation of confidence rankings. The numbers represent different estimates of the probability of the same event  $A$ .

the agent holds with the level of confidence represented by that rank. For example, the judgment that  $\Pr(A) = 0.01$  holds only for the smaller one of the sets represented in figure 2. Therefore an agent whose confidence levels are represented by that figure believes with low confidence that  $\Pr(A) = 0.01$ .

Third, and finally, the higher the confidence level we consider is, the fewer will be the probability judgments that the agent holds with that level of confidence, because fewer judgments are true for all probability measures in larger sets, for example, the judgment that  $\Pr(A) = 0.01$  holds for all the probability measures in the smaller set of the ranking but not in larger sets versus the judgment that  $\Pr(A) \leq 0.05$  may hold at all confidence ranks. The agent holds a given probability judgment with the highest confidence allowed by its rank, namely, with the level corresponding to the larger sets on which that judgment holds. For example, an agent whose confidence levels are represented by figure 2 may believe with high confidence that  $\Pr(A) \leq 0.05$  but have low confidence that  $\Pr(A) = 0.01$ .

Confidence rankings induce an ordinal structure on the set of probability judgments (Hill 2019, 8). More specifically, any elicitation of a complete confidence relation among probability judgments picks up a unique confidence ranking (Bradley, Helgeson, and Hill 2017, 512). Moreover, each complete cardinal ranking of probability judgments can be turned into a confidence ranking, but not vice versa (Hill 2019, 7–9). Therefore the confidence approach can be applied even if the agent does not have enough information to assign a cardinality measure to a probability judgment (Bradley 2017, 297). This is the main difference between the confidence approach and the second-order probability approach that is currently used.<sup>7</sup>

<sup>7</sup> An equivalent representation of confidence levels is due to Bradley (2017). Bradley represents confidence levels as partitions of the set of probability judgments; each set of the relevant partition corresponds to a distinct level of confidence (296–99).



It is also important to clarify the nature of the probabilities at play. In the confidence approach, the objects of confidence are usually first-order beliefs. As Roussos, Bradley, and Frigg (2021, 14) stated, for example, “‘confidence’ is . . . a (second-order) attitude towards a (first-order) claim, reflecting an evaluation of the state of knowledge underpinning it.” In particular, levels of confidence are *credal sets*, that is, sets of credences. However, the confidence approach can be applied in a situation in which the underlying probabilities purport to represent frequencies. In this case, confidence levels are probability sets. For example, Roussos, Bradley, and Frigg apply the approach to insurance pricing by using as input a set of probabilities that are determined by an ensemble of models. This will allow us to apply the same approach to PRA.

### 5 An example: Reading the new Italian Seismic Hazard Model (MPS19)

In this section, I show that the standard approach and the confidence approach give different results in concrete cases. To do this, I consider an example from the new Italian Seismic Hazard Model (MPS19). MPS19 is soon to supersede the old hazard map (MPS04) that has been used for defining seismic zones in Italy since 2006. The new map was commissioned by the Italian government to the Seismic Hazard Centre (CPS) of the Italian Natural Institute of Geophysics and Volcanology in 2015. The models used in MPS19 were solicited directly by the promoters through an open call to the scientific community (Meletti et al. 2021). In the end, MPS19 included eleven different earthquake rate models of the entire Italian territory: five area-based models (M1–M5), two zone-free models (MF1 and MF2), two physical models (MS1 and MS2), and two geodetically based statistical models (MG1 and MG2).

The input data that were used for evaluating the models consisted of a collection of new and existing data sets. These data sets were heterogeneous and included, among other sources, two instrumental catalogs of the Italian region, a database of seismogenic sources, and two assessments of focal mechanisms of expected earthquakes. It was not required that this data set be used to build the models. Only four out of ten data sets were used as input in more than one model, and none was used in all models. One data set was used specifically in the testing phase, and another was used to select and test the ground-motion acceleration models used in MPS19 (Lanzano et al. 2020).

This selection started from a list of more than three hundred published equations and consisted of three phases: (1) fourteen equations were preselected based on their applicability to the Italian territory; (2) those equations were tested for accuracy against a data set of historical recordings and received scores based on their performances; and (3) a final score was assigned to each equation by combining their forecasting skills and the weights assigned by a group of experts, and the three higher-scored equations were selected to enter the final hazard calculations in MPS19. Finally, each model underwent a double weighting procedure. First, the model was included in MPS19 only if it passed a consistency test that assessed if that model was compatible with the catalog data and if it was deemed physically plausible by an independent group of experts; any model that did not pass the consistency test was excluded.<sup>8</sup> Second, a weight  $W = W_1 \cdot W_2 / C$  was assigned to each model, where  $W_1$  is a

<sup>8</sup> In practice, the territory was divided into seven macro-areas for the testing phase of MPS19; if a model did not pass the consistency test in one subregion, its estimates in that subregion were replaced by the weighted average of the outputs of the consistent models.

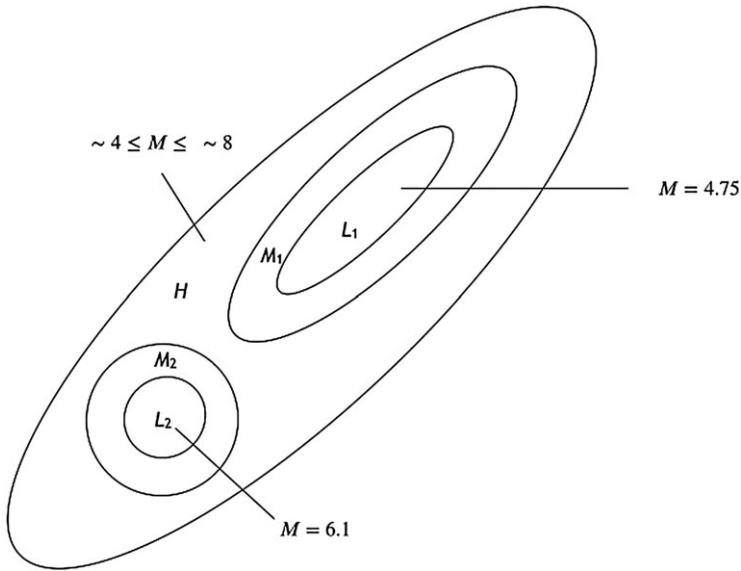
weight based on the testing phase,  $W_2$  is a weight based on experts' judgments, and  $C$  is a normalization factor equal to the sum of  $W_1 \cdot W_2$  for all models. Epistemic uncertainty is represented in MPS19 by reporting a beta distribution whose parameters correspond to the weighted standard deviation of all the hazard estimates and the weighted average value among all estimates. More specifically, the hazard is expressed, not by a single number (the mean hazard), but rather by a distribution the dispersion of which corresponds to the scientific uncertainty concerning the "true" frequency (Meletti et al. 2021, 13–14).

The example that I consider is a single quantity that can be estimated using the MPS19, namely, the frequency of an earthquake with magnitude  $M4.5$  in the Italian region (see Meletti et al. 2021, 10–12). Based on the eleven earthquake-rate models that were included in MPS19, the mean frequency of an earthquake with magnitude  $M4.5$  in the Italian region is between  $\sim 4$  and  $\sim 8$  earthquakes per year (ea/yr). The lowest estimate is  $\sim 3$  eq/yr (from one of the fault-based models), whereas the highest estimated frequency is  $\sim 17$  eq/yr (from one area-source model). The weighted average of the ensemble is  $\sim 5.7$  eq/yr, which is the frequency of  $M4.5$  is  $\sim 5.7$  eq/yr if we adopt the approach described in section 2. By contrast, if we adopt the MPS19 representation of epistemic uncertainty, this frequency will correspond to the interval  $[\sim 4, \sim 8]$  and a probability distribution over that interval. This distribution will show that the overall weight of the models that estimate that frequency to be less than  $\sim 5.7$  eq/yr is 0.38, whereas the overall weight of the models that estimate it to be higher than  $\sim 5.7$  eq/yr is 0.62.

Suppose that we want instead to represent the confidence of the panel of experts with respect to the same quantity (the frequency of an earthquake with magnitude  $M4.5$ ) without using second-order probabilities. We can proceed as follows. The two models that receive the highest scores from the panel of experts are a seismotectonic model, MA4 ( $W_2 = 0.2240$ ), and a smoothed seismicity model, MS1 ( $W_2 = 0.2193$ ). Even though we will no longer assign second-order probabilities to the models, let's assume that the experts would place equal confidence in the estimates of MA4 and MS1 for the same quantities. In particular, the mean frequency of MS1 for an earthquake with  $M4.5$  is  $\sim 4.75$  eq/yr, and the mean frequency of MA4 for the same intensity is  $\sim 6.1$  eq/yr. An expert who is confident that either MA4 or MS1 is correct should consequently hold with high confidence that the frequency of  $M4.5$  is either  $\sim 4.75$  eq/yr or  $\sim 6.1$  eq/yr, with low confidence in values far off from both those estimates. The epistemic uncertainty of the panel of experts can be represented using a confidence ranking with multiple foci, as shown in figure 3.

In this case, experts would not be required to assess their confidence in the models but only in the final estimate of a quantity of interest (in this case, the frequency of an earthquake with magnitude  $M4.5$ ). These judgments need not be precise. For example, the experts may believe with high confidence that the frequency is  $\sim 4.75$  eq/yr without assigning a probability to this frequency being exactly 4.75 eq/yr. The final ranking provides a clearer representation of the epistemic uncertainty than the distribution of second-order probabilities used in MPS19. Figure 3 clearly shows that the panel of experts split their confidence among two frequency intervals.

Moreover, the experts may fail to realize how different values of the parameters in the models contribute to the final hazard. As, for example, Bommer and Scherbaum (2008, 1000) remarked, "before seeing the results of a full hazard disaggregation



**Figure 3.** Confidence ranking with two foci for the frequency  $F$  of an earthquake with  $M_{4.5}$  based on MPS19.  $H$  is high confidence,  $M$  is medium confidence, and  $L$  is low confidence.

[of the relative contributions to that hazard from the range of values of magnitude and distance], [the expert] does not really know what [these values correspond] to in terms of magnitude-distance distribution or ground motion.” For example, experts may believe that  $MA_4$  and  $MS_1$  are plausible overall but find their estimates for  $M_{4.5}$  to be unconvincing. The confidence approach allows us to assess how much the experts trust the final estimate rather than determining their degree of confidence that the model is the best one available.

Finally, figure 3 represents the confidence of the panel of experts; a single expert in the panel may believe that only one of the two models is correct and therefore place high confidence in a single interval (notice that the representation would be the same if half of the panel believes that  $MA_1$  is the best model and the other half believes that  $MS_2$  is the best model). If the confidence levels of the panel are represented by figure 3, the panel believes with high confidence that the frequency of  $M_{4.5}$  is either  $\sim 4.75$  eq/yr or  $\sim 6.1$  eq/yr (because this judgment holds for all the frequencies that are held with high confidence) but does not believe that this frequency is, say,  $\sim 5.4$  eq/yr (because this judgment is false for at least some of the frequencies in the high confidence rank).

## 6 Confidence in probabilistic risk assessment

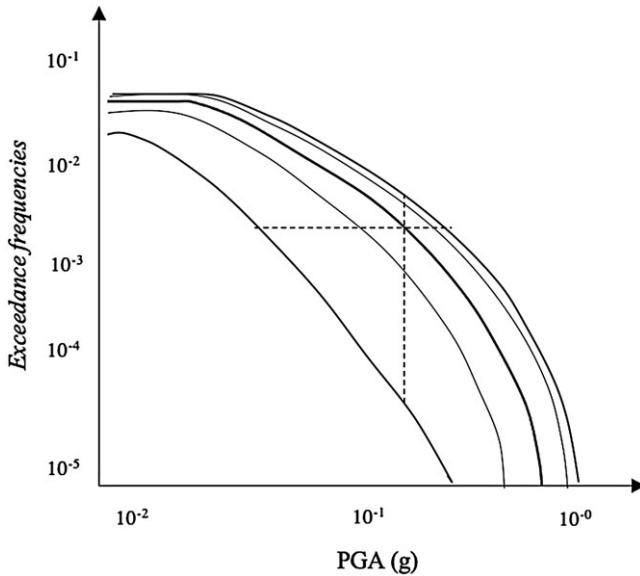
In this section, I show how all this can be turned into a formal representation of epistemic uncertainty in PRA. The confidence approach was originally proposed to represent uncertainty about competing scientific estimates (Hill 2013, 675–76). Roussos, Bradley, and Frigg (2021) extend it to the pricing of hurricane insurance. Given that the models employed in hurricane hazard analysis are different from the

ones of PHA, it will be instructive on its own to see how the same approach can be extended to risk analysis. This will require also two steps: (1) defining a probability set containing all the estimated frequencies of the same event and (2) formulating a confidence ranking over this set. In the next section, I discuss how this approach compares to the standard approach.

The credal set is defined by the ensemble of models. A model whose parameters have all been set to some specific value determines a *hazard curve* that plots intensity levels (on the *x*-axis) and their frequencies of exceedance (on the *y*-axis). The characteristic shape of the hazard curve displays the fact that events that exceed lower intensity thresholds have a higher frequency (top left of the curve), whereas events that exceed higher thresholds have a lower frequency (bottom right of the curve). The ensemble of models can be represented as a family or bundle of hazard curves. This representation conveys two pieces of information. Horizontally, the values on the *y*-axis corresponding to the same point on the *x*-axis represent the variance in the estimation of the frequency of an event with a given intensity. Vertically, the values on the *x*-axis that correspond to the same point on the *y*-axis represent the variance in the estimate of the most intense event with a given frequency.

Both these dimensions are relevant in the practice of risk analysis. For example, a failure rate of 10 percent in fifty years is often considered the safety threshold in earthquake engineering. In this case, we consider the horizontal dimension, whose spread corresponds to the divergence of the models in their estimates of the strongest event that may occur at the site with the relevant frequency. I focus, however, on the vertical dimension, whose spread indicates the variance in the frequency of the event according to the models included in the ensemble. The vertical spread of the bundle of hazard curves can be taken to represent the relevant probability interval that the experts consider for estimating the frequency of exceedance corresponding to a specific intensity. This interval comprises all the probability measures for the same event between the frequency estimated by the “lowest” curve and the frequency estimated by the “highest” curve. A simplified bundle of hazard curves with a corresponding confidence ranking is shown in figure 4. The probability set can be built by including all the frequencies estimated by the models. Alternatively, one can consider the interval between the lowest and highest frequencies. An advantage of using the first method is that it includes only the frequencies that are correct according to some models. At the same time, this overlooks the fact that such estimates have an epistemic uncertainty attached to them. To avoid this problem, the set of frequencies may be represented as a set of intervals (see later).

The panel of experts would then provide a confidence ranking over this probability set. For example, a panel of experts whose confidence levels can be represented as in figure 4 believes with medium confidence that the correct frequency of a specified level of intensity lies between the 16th and the 84th percentiles estimated by the ensemble, while being less confident in any more precise interval. A different expert may be highly confident that the correct frequency corresponds to the mean value and have no confidence that this frequency is anywhere far off the mean. Intuitively, the confidence ranking should be centered around the frequency estimated by the model with the highest weight (e.g., Hill, [forthcoming](#), 7). In this case, the confidence ranking can be represented as in figure 2. However, experts may also believe with



**Figure 4.** A simplified bundle of hazard curves. The curves represent (from the top) the highest curve estimated by the ensemble, the 84th percentile, the mean, the 16th percentile, and the lowest estimated curve. The dotted lines represent the horizontal dimension of the bundle (i.e., the variance in the estimate of the frequency of an event with a given intensity) and its vertical dimension (i.e., the variance in the estimate of the most intense event with a given frequency).

equal confidence that either of two different estimates corresponds to the correct frequency. In this case, the two intervals corresponding to their confidence judgments may be partially overlapping or disjoint.

If these two intervals partially overlap, the frequencies that lie in the intersection of these intervals can be considered more robust estimates than the ones that are correct according to only one model (Roussos, Bradley, and Frigg 2021, 27–34). We can distinguish between two routes to robustness. The first one is given by the standard approach using the logic tree. As seen, each branch of the logic tree corresponds to a hazard curve whose weight is determined by the weights assigned to the segments of that branch. If two models agree on the final estimate of the hazard, then the total probability of this estimate is given by the sum of the weights of the two models. Another route to robustness is as follows. Each model can be represented as a subset of the interval between the maximum and minimum values estimated by the ensemble. If different subsets overlap, the analyst can decide to put the values in the intersection in the confidence ranking so that robust estimates are retained as beliefs (Roussos, Bradley, and Frigg 2021, 28–29; Hill, *forthcoming*, sec. 7, 35–41).

If the two intervals are disjoint, then there is a choice of how to construe the confidence ranking. Earlier, we considered the full interval between the lowest and the highest estimates produced by the model ensemble as the probability set of reference for the definition of confidence rankings. Alternatively, we could have considered the set of point value estimates produced by the models. This would

neglect the fact that each one of these models has multiple submodels that produce a specific epistemic uncertainty around the mean frequency. In the case of disjoint intervals, we can construe the final confidence ranking by considering the intervals between the highest and the lowest estimates that the agent holds with a given level of confidence. For example, in the case discussed in section 5, one could consider the interval between  $\sim 4.75$  eq/yr and  $\sim 6.1$  eq/yr as the low confidence interval (i.e., the expert has at least medium confidence that the true frequency is between  $\sim 4.75$  and  $\sim 6.1$  eq/yr and low confidence in more precise estimates). However, this would misrepresent the epistemic situation of the panel of experts. Indeed, the experts split their confidence between  $\sim 4.75$  eq/yr and  $\sim 6.1$  eq/yr in the sense that they believe that the correct frequency is close to either one of the two estimates. However, the experts would think that many values in between these two estimates add little to the confidence of the prediction, and therefore these experts may be willing to put them in a higher confidence ranking.

The epistemic uncertainty of the experts is best represented by allowing for confidence rankings with multiple foci, as shown in figure 3. Such rankings are sets that contain disjoint families of sets; the set corresponding to a given level of confidence is formed by taking the union of the sets of estimates that the agent holds with that level of confidence. The fact that the confidence ranking has multiple foci does not mean that the expert thinks that there are multiple correct frequencies of the same event but only that the evidence available is such that it permits holding two disjoint probability intervals with equal confidence. This is likely to be the case in the forecasting of rare and extreme events. In particular, the sets  $M_1 \cup M_2$  and  $L_1 \cup L_2$  are not convex. Convexity is a condition in Hill (2013, 681, Definition 2) but not in Hill (2019, forthcoming).<sup>9</sup> By contrast, Roussos (2020, 249) emphasizes that the convexity of confidence rankings is a nonessential feature of the approach, because “the sets in the nesting can contain only the point valued outputs of the models, or balls around those outputs (representing initial condition and parameter uncertainty), without including all of the points between those outputs.” Roussos reserves the term “gappy sets” for confidence rankings that contain multiple intervals of values (250, figure 4). It is plausible that at least in some cases, the uncertainty of the experts is represented by a set with multiple foci; “gappy” sets are therefore apt to represent model uncertainty in PRA.

## 7 Representing model uncertainty

The main differences between this approach and the standard approach in PRA are as follows. First of all, the standard (Bayesian) approach yields a single estimate of hazard (either the mean or a percentile). The mean hazard, which is the reference value in civil and environmental engineering, includes epistemic uncertainty in the sense that it is the weighted average of the estimates produced by models. We also mentioned the MPS19 representation of epistemic uncertainty, in which this uncertainty is represented by a distribution over an interval of frequencies for the same event of interest. By contrast, the approach presented in this article yields

<sup>9</sup> I am grateful to an anonymous reviewer for pointing this out to me.

different sets of estimates, each held by the experts with a different level of confidence.

The representation of epistemic uncertainty used in MPS19, in particular, builds more information in the model about the experts' levels of confidence than licensed by the experts' judgments themselves. To assign a weight to each model, the experts' judgments are first expressed as probabilities, then the weights are assigned to the models on a particular distribution that represents the overall scientific community and is inferred from the judgments expressed by the experts, assuming that those experts are representative of the scientific community. Treating both the experts' elicitation weights and the performance weights of the models as probabilities makes those weights comparable. However, as the MPS19 Working Group acknowledged, "the procedure carries on an unavoidable subjectivity in merging the different scoring metrics of the testing phase and the scoring of the experts' judgement" (Marzocchi and Jordan 2017, 15).

At the same time, the model may also provide too much information for most practical uses. One of the goals of MPS19 is to convey the full range of epistemic uncertainty to the user (stakeholders and decision makers). To do this, it estimates epistemic uncertainty as a probability distribution rather than providing a single value for seismic hazard. As the MPS19 Working Group emphasizes, "MPS19 shows the aleatory variability and the epistemic uncertainty in the final output, making the interpretation of the dispersion of the hazard curves more clear, and providing the interpretation for any possible end-user" (Meletti et al. 2021, 24). However, in most circumstances, end users may *not* need the full range of estimates but only those estimates that are held with some confidence by the scientific community, which may, in turn, depend on what is at stake in those circumstances. By contrast, the confidence approach is clearer and easier to use for decision-making. As emphasized especially by Hill, the confidence in a probability judgment required for that judgment to play a role in a decision depends on what is at stake in that decision. A key feature of this approach is that it allows decision makers to consider only the set of estimates that correspond to the level of confidence that they want to achieve.

Second, the standard approach represents confidence as a cardinal notion (assigning weights to models), whereas the confidence approach represents it as an ordinal notion (using sets of probabilities). Consequently, the confidence approach does not require experts to have enough evidence to warrant precise judgments (Bradley 2017, 297; Roussos, Bradley, and Frigg 2021, 12–13). This approach is therefore particularly justified when the historical data are limited or ambiguous. Moreover, the approach that I present in this article is continuous with the representation of uncertainty that is currently used in climate modeling.

Third, the standard approach is "bottom-up": the experts are tasked with evaluating the components of the models, and the weight of the final estimate produced by a model is calculated by multiplying the probabilities in the logic tree. By contrast, the confidence approach is "top-down": experts assess directly their confidence about the estimate of a single quantity (e.g., the frequency of an event with a specific intensity). More specifically, the inputs of a PRA consist of some sets of historical data from the area in which the models will be applied, and the experts are asked to evaluate the models based on these data. The output then consists of a single hazard estimate that incorporates the epistemic uncertainty of the panel (SSHAC

1997, 177). By contrast, both the empirical data and the array of available models should be provided to the experts as inputs for their evaluation, and the experts should assess their confidence concerning a single quantity of interest. This ranking should reflect the “weight of evidence” available to the experts that is relevant to their judgments (Bradley 2017, 290; Bradley and Hill 2019, 506–7). This evidence may include not only historical recordings but also the estimates produced by different models. Importantly, the amount of evidence supporting a forecast is distinct from the forecast itself. For example, new data about frequent but low-intensity events can make the scientist more confident that a particular model is correct even though her estimates remain the same. In this sense, “confidence [reflects] an evaluation of the state of knowledge underpinning [a probabilistic judgment]” (Roussos, Bradley, and Frigg 2021, 4).

It is important to notice, however, that the two approaches are not in contrast with each other; on the contrary, the confidence framework allows or supports a top-down approach while also being compatible with a bottom-up approach.<sup>10</sup> Helgeson, Bradley, and Hill (2018) emphasize that an advantage of their approach is that it allows for a “downstream transmission” of confidence. In their words, “by building confidence assessments into a formal belief representation (the nested sets), we facilitate this propagation [to ‘downstream’ modeling and decision analysis]” (521). At the same time, an “upstream” transmission from logic trees to the confidence approach is also possible. For example, in MPS19, each model received two weights, one based on the experts’ evaluation and one based on the model’s performance. These performance-based weights can be used to provide a *partial ordering* of the relevant frequencies that can be then passed on to the experts. For example, the ERMs of MPS10 also received a score that was based purely on their forecasting performances with respect to an independent set of data ( $W_1$ ). Two ERMs have  $W_1 = 0.1439$  (MA1 and MS2) and four ERMs have  $W_1 = 0.1136$  (MA4, MF1, MF2, MS1). The mean frequency of MA1 for M4.5 is  $\sim 7$  eq/yr, whereas the mean frequency of MS2 is  $\sim 6.2$  eq/yr. At the same time, the mean frequency of the four models that received the second-best score is between  $\sim 4.5$  and  $\sim 6.3$  eq/yr. If one were to use the performance-based weights to form a confidence ranking, the set corresponding to low confidence would contain  $\sim 6.2$  and  $\sim 7$ , whereas values as low as  $\sim 4.5$  would be included at the next confidence level up.

Finally, and most importantly, in the confidence approach, there is a clear distinction between the models’ hazard estimates and the experts’ confidence judgments. As both Hill (2013, esp. 678–80) and Bradley (2017, sec. 13.4, 277–78) frequently emphasize, the confidence approach allows decision makers to consider the estimates that are held with a level of confidence that is appropriate for a specific situation. Once the appropriate set is selected, the confidence approach is compatible with different decision rules. Many such rules have been proposed. For example, Hill (2019, 9–11) mentions the maximin expected utility rule, which requires choosing the option with the greatest minimum expected utility. Hill (forthcoming, 37) argues that “whilst the [decision maker] may act as a subjective expected utility agent with the aggregate (centre) probability when the stakes are low, when the stakes are higher, it

<sup>10</sup> I owe the point and this formulation to an anonymous reviewer.



will use a larger set of priors, and . . . be more ambiguity averse.” It is worth noting that this flexibility with respect to decision rules is not available in the standard approach. By contrast, in PRA, the hazard is *defined* as the weighted average of the estimates produced by the models. The standard approach therefore includes linear pooling in the definition of the hazard. By contrast, the confidence approach represents the models’ estimates and the experts’ judgments separately and makes both available in the decision.

## 8 Conclusions

Epistemic uncertainties are included in PRA as second-order probabilities that represent the degree of belief of the analyst that a model is correct. However, experts typically do not have enough evidence to warrant precise judgments. In this article, I presented a methodology that incorporates the analyst’s confidence in a given probability interval. My proposal extends Hill’s and Bradley’s *confidence approach*. Following this approach, model uncertainty is represented using nested families of probability sets with multiple foci. The main differences between the standard approach and the confidence approach are as follows. (1) The standard (Bayesian) approach yields a single estimate of hazard (either the mean or a percentile), whereas the approach presented in this article yields different sets of estimates held with a particular level of confidence. (2) In the standard approach, confidence is a cardinal notion (weights in the logic tree), whereas the confidence approach represents it as an ordinal notion (confidence rankings). (3) The standard approach is “bottom-up,” whereas the confidence approach is “top-down.” (4) In the confidence approach, there is a clear distinction between the models’ estimates and the experts’ confidence. The confidence approach is, therefore, clearer, less questionable, and more usable in decisions than the Bayesian approach.

**Acknowledgments.** This research was supported by MUR–Italian Ministry of University and Research, PRIN Scheme (“Understanding Scientific Disagreement and Its Impact on Society,” project P2022A8F82). This article was written during my visit to the Center for Philosophy of Natural and Social Sciences of the London School of Economics and Political Sciences. I am grateful to Lorenza Petrini and Daniele Chiffi for introducing me to the subtleties of probabilistic risk analysis and severe uncertainty and to Richard Bradley for supporting this project. Parts of this article were presented at the Varieties of Risk Seminar of the University of Edinburgh, at the Centre for Philosophy and the Sciences of the University of Oslo, and at the triennial conference of the Italian Society for Logic and the Philosophy of Science. I am particularly grateful to Philip Ebert, Rafal Urbaniak, Jack Wright, Anders Strand, Roman Frigg, and Malvina Ongaro for feedback and discussion. Finally, I thank two anonymous reviewers for their comments, which definitely improved this article.

## References

- Abrahamson, N., and J. Bommer. 2005. “Probability and Uncertainty in Seismic Hazard Analysis.” *Earthquake Spectra* 21 (2):603–7. <https://doi.org/10.1193/1.1899158>.
- Aven, T., and O. Renn. 2015. “An Evaluation of the Treatment of Risk and Uncertainties in the IPCC Reports on Climate Change.” *Risk Analysis* 35 (4):701–12. <https://doi.org/10.1111/risa.12298>.
- Baker, J., B. Bradley, and P. Stafford. 2021. *Seismic Hazard and Risk Analysis*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108425056>.
- Bommer, J. J. 2012. “Challenges of Building Logic Trees for Probabilistic Seismic Hazard Analysis.” *Earthquake Spectra* 28 (4):1723–35. <https://doi.org/10.1193/1.4000079>.

- Bommer, J. J., and F. Scherbaum. 2008. "The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis." *Earthquake Spectra* 4 (24):997–1009. <https://doi.org/10.1193/1.2977755>.
- Bradley, R. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9780511760105>.
- Bradley, R., C. Helgeson, and B. Hill. 2017. "Climate Change Assessments: Confidence, Probability and Decision." *Philosophy of Science* 84 (3):500–522. <https://doi.org/10.1086/692145>.
- Budnitz, R., G. Apostolakis, D. Boore, L. Cluff, K. Coppersmith, C. Cornell, and P. Morris. 1998. "Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis." *Risk Analysis* 18 (4):463–69. <https://doi.org/10.1111/j.1539-6924.1998.tb00361.x>.
- Helgeson, C., R. Bradley, and B. Hill. 2018. "Combining Probability with Qualitative Degree-of-Certainty Metrics in Assessment." *Climatic Change* 149:517–25. <https://doi.org/10.1007/s10584-018-2247-6>.
- Hill, B. 2013. "Confidence and Decision." *Games and Economic Behavior* 82:675–92. <https://doi.org/10.1016/j.geb.2013.09.009>.
- Hill, B. 2019. "Confidence in Beliefs and Rational Decision Making." *Economics and Philosophy* 35 (2):223–58. <https://doi.org/10.1017/S0266267118000214>.
- Hill, B. Forthcoming. "Confidence, Consensus and Aggregation." HEC Paris Research Paper. <https://ssrn.com/abstract=4387641>.
- Janzwood, S. 2020. "Confident, Likely, or Both? The Implementation of the Uncertainty Language Framework in IPCC Special Reports." *Climatic Change* 162:1655–75. <https://doi.org/10.1007/s10584-020-02746-x>.
- Kammerer, Annie M., and Jon P. Ake. 2010. *Implementation Guidance for SSHAC Level 3 and 4 Processes*. Report NUREG-2117. Washington, DC: US Nuclear Regulatory Commission. <https://www.nrc.gov/docs/ML1207/ML12073A311.pdf>.
- Klügel, J. 2011. "Uncertainty Analysis and Expert Judgement in Seismic Hazard Analysis." *Pure and Applied Geophysics* 168:27–53. <https://doi.org/10.1007/s00024-010-0155-4>.
- Kulkarni, R. B., R. R. Youngs, and K. J. Coppersmith. 1984. "Assessment of Confidence Intervals for Results of Seismic Hazard Analysis." *Proceedings, Eighth World Conference on Earthquake Engineering* 1:263–70.
- Lanzano, G., L. Luzi, V. D'Amico, F. Pacor, C. Meletti, M. Marzocchi, R. Rotondi, and E. Varini. 2020. "Ground Motion Models for the New Seismic Hazard Model of Italy (MPS19): Selection for Active Shallow Crustal Regions and Subduction Zones." *Bulletin of Earthquake Engineering* 18:3487–516. <https://doi.org/10.1007/s10518-020-00850-y>.
- Marzocchi, W., and T. Jordan. 2017. "A Unified Probabilistic Framework for Seismic Hazard Analysis." *Bulletin of the Seismological Society of America* 107 (6):2738–44. <https://doi.org/10.1785/0120170008>.
- Mastrandrea, M., K. Mach, G. Plattner, O. Edenhofer, T. Stocker, C. Field, K. Ebi, and P. Matschoss. 2011. "The IPCC AR5 Guidance Note on Consistent Treatment of Uncertainties: A Common Approach across the Working Groups." *Climatic Change* 108: Article 675. <https://doi.org/10.1007/s10584-011-0178-6>.
- McGuire, R., C. A. Cornell, and G. Toro. 2005. "The Case for Using Mean Seismic Hazard." *Earthquake Spectra* 21 (3):879–86. <https://doi.org/10.1193/1.1985447>.
- Meletti, C., W. Marzocchi, V. D'Amico, G. Lanzano, L. Luzi, F. Martinelli, B. Pace, A. Rovida, M. Taroni, F. Visini, and the MPS19 Working Group. 2021. "The New Italian Seismic Hazard Model (MPS19)." *Annals of Geophysics* 64 (1). <https://doi.org/10.4401/ag-8579>.
- Musson, R. 2005. "Against Fractiles." *Earthquake Spectra* 21 (3):887–91. <https://doi.org/10.1193/1.1985445>.
- Musson, R. 2012. "On the Nature of Logic Trees in Probabilistic Seismic Hazard Assessment." *Earthquake Spectra* 28 (3):1291–96. <https://doi.org/10.1193/1.4000062>.
- O'Hagan, A., C. Buck, A. Daneshkhan, R. Eiser, P. Garthwaite, D. Jenkinson, J. Oakley, and T. Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. London: John Wiley.
- Reisinger, A., M. Garschagen, K. Mach, M. Pathak, E. Poloczanska, M. van Aalst, A. Ruane et al. 2020. "The Concept of Risk in the IPCC Sixth Assessment Report: A Summary of Cross-Working Group Discussions: Guidance for IPCC Authors." <https://www.ipcc.ch/event/guidance-note-concept-of-risk-in-the-6ar-cross-wg-discussions/>.
- Roussos, J. 2020. "Policy-making under Scientific Uncertainty." PhD thesis, London School of Economics and Political Science. <http://etheses.lse.ac.uk/4158/>.
- Roussos, J., R. Bradley, and R. Frigg. 2021. "Making Confident Decisions with Model Ensembles." *Philosophy of Science* 88 (3):439–60. <https://doi.org/10.1086/712818>.

- Scherbaum, F., and N. M. Kuehn. 2011. "Logic Tree Branch Weights and Probabilities: Summing Up to One Is Not Enough." *Earthquake Spectra* 27 (4):1237–51. <https://doi.org/10.1193/1.3652744>.
- Secanell, R., C. Martin, E. Viallet, and G. Senfaute. 2018. "A Bayesian Methodology to Update the Probabilistic Seismic Hazard Assessment." *Bulletin of Earthquake Engineering* 16:2513–27. <https://doi.org/10.1007/s10518-017-0137-3>.
- Selva, J., and L. Sandri. 2013. "Probabilistic Seismic Hazard Assessment: Combining Cornell-like Approaches and Data at Sites through Bayesian Inference." *Bulletin of the Seismological Society of America* 103 (3):1709–22. <https://doi.org/10.1785/0120120091>.
- Senior Seismic Hazard Analysis Committee. 1997. *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*. Report NUREG-CR-6372. Washington, DC: US Nuclear Regulatory Commission. <https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6372/vol1/index.html>.
- Wüthrich, N. 2016. "Conceptualizing Uncertainty: An Assessment of the Uncertainty Framework of the Intergovernmental Panel on Climate Change." In *Recent Developments in the Philosophy of Science: EPSA15*, edited by M. Massimi, J. Romeijn, and G. Schurz, 95–107. Dusseldorf, Germany: Springer. [https://doi.org/10.1007/978-3-319-53730-6\\_9](https://doi.org/10.1007/978-3-319-53730-6_9).