Chapter 19
Representing Uncertainty

Sven Ove Hansson

Abstract Our uncertainty about matters of fact can often be adequately represented by probabilities, but there are also cases in which we, intuitively speaking, know too little even to assign meaningful probabilities. In many of these cases, other formal representations can be used to capture some of the prominent features of our uncertainty. This is a non-technical overview of some of these representations, including probability intervals, belief functions, fuzzy sets, credal sets, weighted credal sets, and second order probabilities.

19.1 Uncertainty in Decisions

Many decisions are difficult because we do not know the effects of our alternatives. Consider the following examples:

– You have been offered a free two-day hang-gliding course, but you partner is worried, and says: "You must first find out how dangerous it is."
– You are tempted to buy a lottery ticket. The top prize would solve all your financial problems, but the ticket is quite expensive. Should you buy it?
– You are offered to bet on a horse but you do not know its chances.
– The authorities have to decide if a new chemical can be used, but they do not know what effects it may have on human health and the environment.

In the first two cases it is likely that probabilities can be of some help (but the decision may still be a very difficult one). In the last two cases it is not obvious that probabilities can at all be used. Can formal methods nevertheless be used to throw light on decision problems like these?

The formal representation of uncertainty has mostly been discussed in contexts of decision-guidance, but the topic is interesting also apart from applications to decision-making. Formal representations can be used to distinguish between

S. O. Hansson (✉)
Division of Philosophy, Royal Institute of Technology (KTH), Stockholm, Sweden
e-mail: soh@kth.se

S. O. Hansson, V. F. Hendricks (eds.), *Introduction to Formal Philosophy*, Springer Undergraduate Texts in Philosophy, https://doi.org/10.1007/978-3-319-77434-3_19

different kinds of ignorance or lack of knowledge, and they can also contribute to our understanding of phenomena such as belief change and conditional beliefs.

19.2 Possibility Sets and Weight Functions

You can be uncertain about various matters, for instance about the state of the world, how to describe it, what options are available to you in a decision, your evaluation of various outcomes, and your own moral and philosophical principles. This chapter is devoted to uncertainty about the state of the world and about the correctness of our descriptions of it.[1] The simplest representation of such uncertainty is what we may call a *possibility set*, a set consisting of the alternatives one is uncertain between. Suppose that I know that either Ann, Bob, or Cai baked the cake, but I do not know which of them. Then the cake was baked by a member of the possibility set {Ann, Bob, Cai}.

In many cases, our uncertainty concerns a number, i.e. the value of a numerical variable. Then the possibility set will be a set of numbers, a *numerical possibility set*. In the most common applications, such sets have the form of intervals. Even if you do not know how much money you have on your account, you may know that you have between €500 and €1000. Then the number of euros on your account is an element of [500, 1000], the set of real numbers not lower than 500 and not higher than 1000. The common rules of arithmetic can be extended to intervals, as outlined in Box 19.1 and illustrated in the following examples [22, pp. 11–14]:

> She has between €1000 and €2000 and he has between €500 and €700. Thus, together they have between €1500 and €2700, and she has between €300 and €1500 more than he has.
>
> There will be between 10 and 20 participants in the competition, and each of them will use between 3 and 5 fishing-rods. Therefore, between 30 and 100 fishing-rods will be used.

Box 19.1 Some rules of interval arithmetic

$$[a, b] + [c, d] = [a + c, b + d]$$
$$[a, b] - [c, d] = [a - d, b - c]$$
$$[a, b] \times [c, d] = [\min(a \times c, a \times d, b \times c, b \times d), \max(a \times c, a \times d, b \times c, b \times d)]$$
$[a, b] \times [c, d] = [a \times c, b \times d]$ if a, b, c, and d are all non-negative.

[1] It is sometimes unclear whether an agent's uncertainty in a particular matter concerning herself is attributable to her lack of factual information or to the fact that she has not made some decision that could have resolved the uncertainty. This type of "ambiguous" uncertainty underlies several of the well-known decision-theoretical paradoxes [15].

Possibility sets provide a simple but also very limited form of uncertainty representation. We often have reasons to differentiate between the elements, and put more weight on some than on the others. If I know that Cai is very fond of baking, then I may give more weight to the possibility of her being the baker than to Ann or Bob. This can be expressed by assigning numbers to each of them. I may for instance introduce a function f such that $f(\text{Ann}) = 0.25$, $f(\text{Bob}) = 0.25$, and $f(\text{Cai}) = 0.50$. This is our second major form of uncertainty representation, a *weighted possibility set*. It consists of a possibility set and a *weight function* that assigns a weight to each of its elements. We will assume that the weights are non-negative and that they add up to 1.

The weight function is often construed as a probability function. It is important to recognize that the notion of probability has a precise mathematical definition. A probability function is a function that satisfies the laws of probability, the laws obeyed by random events in the real world, such as throws of dice and coins. These laws are elegantly summarized in the Kolmogorov axioms that are given in Box 19.2. A mathematical entity that does not satisfy these laws should not be called "probability".

There are two major interpretations of probabilities. First, as in Fig. 19.1, we can think of them as mental entities. In that case, if you assign the probability 1/6 to the dice yielding a six, you make a report on your own state of mind, a "subjective" probability. Alternatively, as in Fig. 19.2, we can think of probabilities as properties of the physical world, existing independently of our minds. In that case your report on the dice is a statement about (tendencies in) the world, an "objective" probability. These two interpretations lead us to quite different developments in the representation of uncertainties. Let us begin with the former.

19.3 The Subjective Interpretation

Many proponents of the subjective interpretation are Bayesians. They maintain that in order to be rational, a person's subjective degrees of belief have to comply with the probability axioms. According to this view, rational uncertainty can always be represented by a probability function. If you are uncertain about whether Bern is the capital of Switzerland, then there must be some definite probability lower than 1 that you can assign to the statement "Bern is the capital of Switzerland". For the Bayesian, uncertainty is just another name of probability.

The major argument for this position is also an argument for the maximization of expected (i.e. probability-weighted) utility. It can be shown that a person who does not abide by this decision rule, or whose probability assignments violate the probability axioms, can have a Dutch book made against her. By this is meant a bet that she would be sure to lose whatever the outcome of the game would be. (See Box 19.3.) Since this is irrational behaviour, it can then be concluded that in order to be rational we should all make our decisions in accordance with a subjective probability assignment that obeys the axioms [16, pp. 381–382].

Box 19.2 Axioms for degree-of-belief representations

The Kolmogorov axioms for probability

Let Ω be the set of all events under consideration. $A, B, \ldots$ are sets of events, i.e. subsets of Ω. Let p be a function from sets of events to real numbers. Then p is a *probability function* if and only if it satisfies the following three axioms:

1. $p(A) \geq 0$ for all $A \subseteq \Omega$. (non-negativity)
2. $p(\Omega) = 1$ (normalization)
3. If $A \cap B = \varnothing$, then $p(A \cup B) = p(A) + p(B)$. (finite additivity)

The above formulation can only be used if the number of events (elements of Ω) is finite. In the more general case, axiom 3 has to be reformulated, and the following should hold for an infinite series $A_1, A_2, \ldots$ of events such that no two of them have an element in common:

3'. $p(A_1 \cup A_2 \cup \ldots) = \sum_{n=1}^{\infty} A_n$ (additivity)

Axioms for other degree-of-belief representations

Alternative degree-of-belief representations usually satisfy the first two but not the third of the Kolmogorov axioms. Instead of (3), the belief function Bel of Dempster-Shafer theory satisfies the following:

4. If $A \cap B = \varnothing$, then $\mathsf{Bel}(A \cup B) \geq \mathsf{Bel}(A) + \mathsf{Bel}(B)$. (finite superadditivity)

In possibility theory, (3) is replaced by the following requirement on the possibility measure Poss:

5. $\mathsf{Poss}(A \cup B) = \mathfrak{max}(\mathsf{Poss}(A), \mathsf{Poss}(B))$

Importantly, this argument refers to the requirements of rational decision-making, not those of rational belief. The demands of rational action (practical rationality) and those of rational belief (theoretical rationality) need not coincide. Holding certain beliefs may be rational in the sense of furthering the achievement of practical goals without being rational in the ratiocinative sense. The Dutch book argument does not tell us what we should believe, only what we should act as believing.

Apart from that, empirical studies have shown that human behaviour commonly violates the Kolmogorov axioms [5, 10]. Therefore, if we are looking for a representation of actual human uncertainty, then we may have good reasons to look for one that does not satisfy the axioms. And of course, if we are not convinced

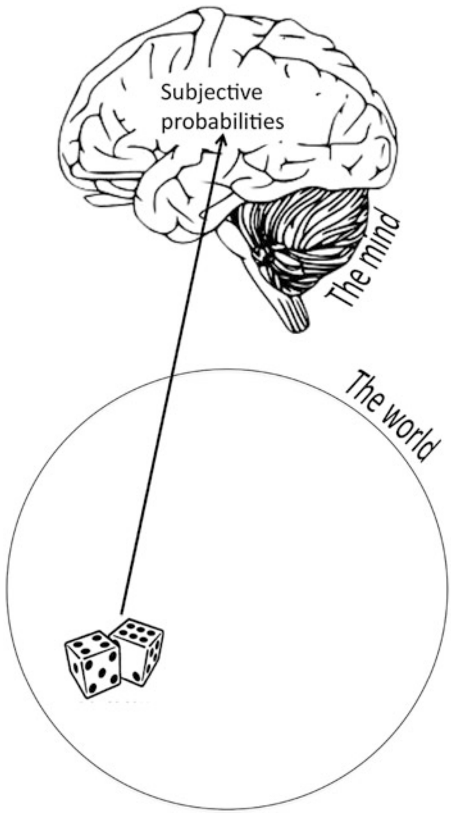

Fig. 19.1 Subjective probabilities

by the Dutch book argument, then we are free to choose rationality principles for (decision-guiding) beliefs based on other criteria.

Several numerical representations of subjective degrees of belief have been presented that are weaker, i.e. require less, than the Kolmogorov axioms. Among the most important of these are the so-called *belief functions* of Dempster-Shafer theory [2, 25]. The characteristic difference between probability functions and belief functions is that the latter can assign more weight to a set than the sum of what it assigns to its elements. For an example, consider the two mutually exclusive events "The Nigerian team will defeat the South African one tomorrow" (N) and "The South African team will defeat the Nigerian team tomorrow" (S). First suppose that your degrees of belief are represented by a probability function p that assigns the value 0.1 to each of the two events, i.e. $p(N) = p(S) = 0.1$. It then follows that $p(\{N, S\}) = 0.2$, i.e. the probability of either team winning is the sum of the probabilities of each of them winning. Next, suppose that instead, your degrees of belief are represented by a belief function Bel, also such that $\mathsf{Bel}(N) = \mathsf{Bel}(S) = 0.1$. From this we cannot conclude what the value of $\mathsf{Bel}(\{N, S\})$ is. It may be equal to 0.2, but it may also be higher, for instance 0.8. This is sensible under some

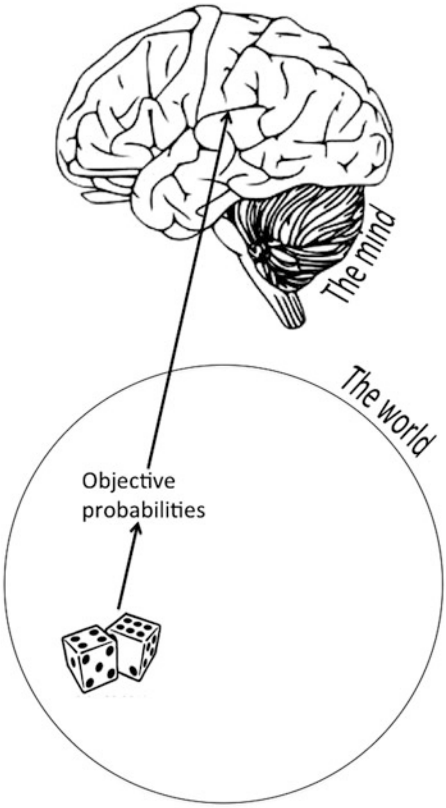

Fig. 19.2 Objective probabilities

interpretations of Bel, since you may have reasons to believe that either *N* or *S* will happen, without having much evidence in favour of either of the two alternatives.

All this has been based on the assumption that the subjective uncertainty that we intend to represent reflects our lack of knowledge about the physical world. However, there are also other sources of uncertainty, in particular vagueness and ambiguities in our language.

ADILAH: Look at him! How would you describe him? Is he bald or not?
MIAHUA: I am inclined to call him bald, but I do not know for sure.

Miahua is not uncertain about what the man's head looks like, but she is uncertain about how to use the word "bald". She may assign weights to the two alternatives, perhaps 0.6 to "bald" and 0.4 to "not bald". However, these numbers need not be interpretable as probabilities. Another option is to interpret them as indicating *fuzzy set* membership.

Box 19.3 An example of a Dutch book

Bob entertains the following beliefs about a dice:

The probability that it yields “1” is 1/6.
The probability that it yields “4” is 1/6.
The probability that it yields “6” is 1/6.
The probability that it yields a prime number is 4/6.

Martha offers Bob the following bets:

If you pay €2 you get €13 if it yields “1”.
If you pay €2 you get €13 if it yields “4”.
If you pay €2 you get €13 if it yields “6”.
If you pay €8 you get €13 if it yields a prime number.

Since Bob believes all these bets to be favourable he accepts them all. As a result of this, he pays €14, and whatever the outcome of the throw he will receive €13 back. This is a sure loss, in other words a Dutch book. The reason for this failure is that his probability assignments do not satisfy the Kolmogorov axioms. (See Box 19.2.)

In common (“crisp”) set theory, there is always a definite answer to whether or not an object is an element of a given set. A set can be represented by an indicator function (membership function, element function). Let μ_A be the indicator function of a set A. Then for all x, $\mu_A(x)$ is equal to 1 if x is an element of A, and equal to 0 if it is not. The function does not take any other value than 0 or 1. In contrast, the indicator function of a fuzzy set can take any value in the interval [0, 1]. If $\mu_A(x) = .5$, then x is “half member” of A, and if $\mu_A(x) = .6$, then it is somewhat more member than non-member [28]. Fuzzy sets can be used to represent vagueness, such as the vagueness that made Miahua uncertain whether the man was bald or not [26, p. 27]. Fuzzy set membership does not satisfy Kolmogorov’s axiom system.

Possibility theory is a variant of fuzzy set theory in which a function Poss with the properties of an indicator function replaces the probability function as a measure of uncertainty [3, 29].

19.4 Credal Sets

Let us return to Fig. 19.2, and to the account of probabilities that treats them as properties of the physical world. This is an interpretation with a strong intuitive appeal. Suppose I tell you that the probability that a certain dice will yield a six is 1/6. You investigate the dice and find that it is loaded and yields a six in about 1 in 3 throws. When telling me this, you expect me to say “Oh, then I was wrong", rather than "But what I said was right, since I reported a probability, and probabilities are states of mind". In cases like these we take probabilities to represent the (counterfactual) frequency that would be recorded if a similar triggering event were repeated under similar circumstances a large number of times. The probability that an atom of Fermium-257 will decay within the next 100 days is close to 0.5. This is a property of the natural world, not merely a subjective belief.

But things do not end here. If there are objective probabilities, then we can be (subjectively) uncertain about which these probabilities are. Consider the following example:

> There are two urns in the room. One of them contains 5 red and 95 black balls. The other contains 95 red and 5 black ones. Someone puts one of the urns − you do not know which − in front of you and asks: "If you draw a ball from this urn, what is the probability that it is red?"

A quite natural answer would be "It is either 0.05 or 0.95". More precisely, you hesitate between two probability functions, p_1 and p_2, which differ in the probabilities they assign to the event (R) that the ball you draw is red. We have $p_1(R) = 0.05$ and $p_2(R) = 0.95$. Your uncertainty can then be expressed with the simplest uncertainty representation introduced above, namely a possibility set $\{p_1, p_2\}$. A possibility set that has probability functions as its elements is called a *credal set* ([20]; cf. [4]). This way of representing uncertainty has been favoured by many philosophers, and also developed in considerable technical detail by statisticians [1]. As one example of its use, climatologists studying the effects of climate change on the probability of extreme weather events employ a range of credible models that generate different probability functions. The outcome of their calculations can then be expressed as a credal set, containing a range of probability functions [9].

In many cases, uncertainty about a specific probability can be expressed as a *probability interval*. Summarizing the outcomes of climate modelling, we may say for instance: "Around 2050, the yearly probability of this whole valley being flooded will be between 1% and 4%." In general, if we have a credal set $\{p_1, \ldots, p_n\}$, then for each event A we can find the minimal probability that it can assign to an event A, i.e. the lowest value of $p_k(A)$ for any p_k. This is the lower probability generated by the credal set, often denoted $\underline{p}(A)$. The upper probability $\overline{p}(A)$ is defined analogously, and together they form the probability interval.

Probability intervals are an intuitively accessible and often quite useful way to express uncertainties [11]. However, it must be observed that information is lost when we simplify a credal set $\{p_1, \ldots, p_n\}$ to the pair $\langle \underline{p}(A), \overline{p}(A) \rangle$ [27]. Consider the following example:

> A dollar coin has been found among the property of a deceased cardsharp. We suspect that it may be unfair. We do not know in which direction it would then be biased, but we know that the most biased coins available to him yield either heads or tails with a frequency of 90%.

Here, the probability of heads in a single throw can be represented by a credal set containing all the probability functions assigning probabilities between 0.1 and

0.9 to heads in a single throw (H). Therefore, $\underline{p}(H) = 0.1$ and $\overline{p}(H) = 0.9$. At first sight it seems as if we can perform further calculations using just $\underline{p}(H)$ and $\overline{p}(H)$. For example, the probability of three heads in a row (HHH) is anywhere in the interval [0.001, 0.729], so that the lower limit coincides with $(\underline{p}(H))^3$ and the upper with $(\overline{p}(H))^3$. However, this pattern cannot be generalized. The probability of getting heads in the first but not in the second throw (HT) is anywhere in the interval [0.09, 0.25]. This interval cannot be calculated from $\underline{p}(H)$ and $\overline{p}(H)$. In fact it is not difficult to construct a credal set that has the same values of $\underline{p}(H)$ and $\overline{p}(H)$ but a different value of $\overline{p}(HT)$.[2] As this example shows, the use of probability intervals instead of the full credal set is as mathematically precarious as it is intuitively accessible, and great care must therefore be exercised in the calculative use of such intervals.

19.5 Weighted Credal Sets

The elements of a credal set may differ in their credibility. To represent these differences we can assign weights to them, thus obtaining a weighted credal set. Most commonly, these weights are probabilities, which leads to a model with probabilities on two levels, as illustrated in Fig. 19.3.

The use of two levels of probabilities (first- and second-order probabilities) has been put in doubt by philosophers since David Hume [17, pp. 182–183] who have worried that if we allow for two levels, then there is no way to avert an infinite regress of higher and higher orders of probability. However, since the two levels in this model represent different types of entities, the process that took us from the first to the second level cannot necessarily be repeated. If it cannot, then no such regress gets started.

Another common criticism is that the two levels of probabilities are unnecessary since they can always be reduced to one. The following variant of our cardsharp example can be used to assess that argument:

> Among the property of the deceased cardsharp we find a dime that may or may not be biased. All biased dimes that have been made have either a 0.1 or a 0.9 probability of heads. Our credal set is $\{p_1, p_2, p_3\}$, where $p_1(H) = 0.1$, $p_2(H) = 0.5$, and $p_3(H) = 0.9$. Furthermore, our subjective probability function $\mathfrak{p}$ over the credal set is such that $\mathfrak{p}(p_1) = 0.25$, $\mathfrak{p}(p_2) = 0.5$, and $\mathfrak{p}(p_3) = 0.25$.

We are now going to throw the dime once. Given all this information, what probability should we assign to it landing heads? The obvious answer is 0.5. We obtain this answer by reducing the two levels to one [23, p. 58]. Letting $\hat{\mathfrak{p}}$ denote the resulting reductive probability function we have:

[2]Let the credal set consist of all probability functions that assign to $p(H)$ a value in either of the two intervals [0.1, 0.2] and [0.8, 0.9]. Then we still have $\underline{p}(H) = 0.1$ and $\overline{p}(H) = 0.9$, but $\overline{p}(HT) = 0.16$.

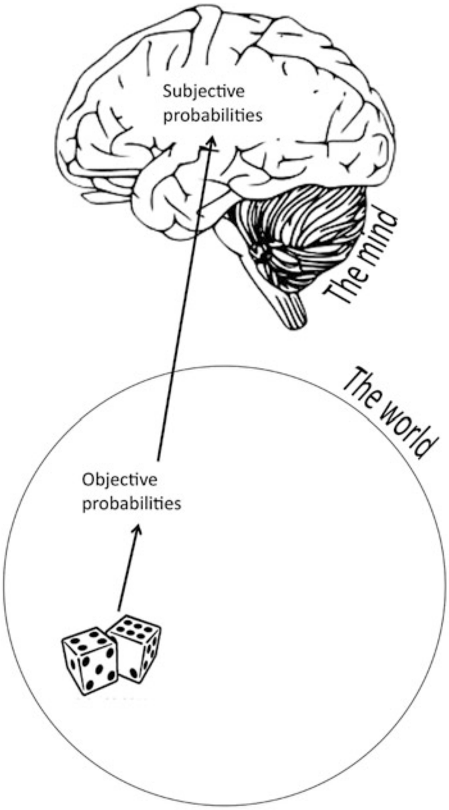

Fig. 19.3 Two-levelled probabilities

$$\hat{\mathfrak{p}}(H) = \mathfrak{p}(p_1) \times p_1(H) + \mathfrak{p}(p_2) \times p_2(H) + \mathfrak{p}(p_3) \times p_3(H) = 0.5$$

At this point, a critic of the two-levelled model might well ask: "So what is the point? Why all this trouble when you could instead, directly, have assigned the subjective probability 0.5 to H?" But there is a good answer to that. Suppose that we also want to know the probability of obtaining three heads in a row when the coin is tossed thrice. That probability is:

$$\hat{\mathfrak{p}}(HHH) = \mathfrak{p}(p_1) \times p_1(HHH) + \mathfrak{p}(p_2) \times p_2(HHH) + \mathfrak{p}(p_3) \times p_3(HHH) = 0.245$$

which is different from $(\hat{\mathfrak{p}}(H))^3 = 0.125$. Thus, if we had followed the advice of the critic we would have obtained the wrong answer to problems involving iterated tosses of the coin. In order to get the right answer in a one-levelled model we would have to give up the assumption that the different tosses in a series are independent events, which is the precondition for deriving the probability of HHH from that of H in the standard way [13]. Hence, the reduction of two-levelled to one-levelled probabilities comes at a high price, namely that we cannot treat events as independent which we intuitively perceive as such. In order to make sense of the complex world that we live in we have to treat some events as similar but independent. This applies to both everyday and scientific reasoning. It is for instance essential for the conceptualization of a repeatable scientific experiment. Although two-levelled probabilities can in principle always be reduced to a single level, the reduced account often has much less explanatory value.

The two-levelled model also has the advantage of providing a lucid account of how empirical information makes us change our views on objective (physical) probabilities. Suppose that we throw the above-mentioned coin five times, and get a series of five heads ($HHHHH$ or in short H^5). This makes us revise $\mathfrak{p}$ and, in particular, increase the value of $\mathfrak{p}(p_3)$, but how much? This we can find out with standard conditionalization on the second level, i.e.

$$\mathfrak{p}(p_3 \mid H^5) = \frac{\mathfrak{p}(H^5 \mid p_3) \times \mathfrak{p}(p_3)}{\mathfrak{p}(H^5 \mid p_1) \times \mathfrak{p}(p_1) + \mathfrak{p}(H^5 \mid p_2) \times \mathfrak{p}(p_2) + \mathfrak{p}(H^5 \mid p_3) \times \mathfrak{p}(p_3)}$$
$$\approx 0.90$$

and similarly $\mathfrak{p}(p_1 \mid H^5) \approx 2 \times 10^{-5}$ and $\mathfrak{p}(p_2 \mid H^5) \approx 0.10$. We now have a revised second-level probability function, $\mathfrak{p}(\ \mid H^5)$, which directly gives rise to a new reductive probability function $\hat{\mathfrak{p}}(\ \mid H^5)$. With its help we can answer the question “Given what we know after observing H^5, what is now our best estimate of the probability that a toss of this coin yields heads?” The answer is $\hat{\mathfrak{p}}(H \mid H^5) \approx 0.86$. This reductive conditionalization is of course very different from standard conditionalization. An ordinary probability function p yields $p(H \mid H^5) = 1$ since given that H^5 took place it is certain that H also took place [13].

The following example illustrates the practical importance of this type of probability revision:

> Two estimates have been made of the probability that a major explosion will take place in the first year’s operation of a new explosives factory (E). According to one estimate, that probability is $p_1(E) = 0.0001$, and according to the other it is $p_2(E) = 0.01$. We believe the former estimate to be much more reliable than the latter, and we have $\mathfrak{p}(p_1) = 0.999$ and $\mathfrak{p}(p_2) = 0.001$. Hence the (reductive) probability of an explosion is $\hat{\mathfrak{p}}(E) \approx 0.00011$, which means that we can almost disregard p_2.
>
> But after a couple of months a major explosion takes place. We therefore update our second-level probability and obtain $\mathfrak{p}(p_1 \mid E) \approx 0.91$ and $\mathfrak{p}(p_2 \mid E) \approx 0.09$. These estimates, rather than the original ones, should be used for instance when we consider whether to open another factory of exactly the same type.

Revisions of this type are an essential mechanism for learning from experience. The change in an epistemic agent’s estimate of the (objective) probability of an event E that ensues after learning that E has occurred can be used as a numerical measure of the agent’s uncertainty concerning the probability of E [14].

We have focused in this section on the use of probability functions on the second level in a two-levelled system. However, proposals have also been made to assign weights other than probabilities to credal sets, such as fuzzy sets or other measures of "epistemic reliability" that do not satisfy the Kolmogorov axioms [7].

19.6 Decision-Making Under Uncertainty

Decision rules for uncertainty can be classified according to the information that they require about the options that are available to be chosen between.[3]

Possibility sets. Suppose that for each of the options that we can choose between we have a possibility set telling us what outcomes can follow, but we know nothing about the comparative plausibility of these outcomes. In this case we have to base the decision on the values of the outcomes. The most common decision rule for this purpose is the *maximin* rule. It tells us to identify for each option its security level, i.e. the value of the worst outcome that it can result in. We should then choose one of the options with the highest security level. According to a variant of this rule, the *leximin rule* (lexicographic maximin), if there is more than one alternative with the highest security level, then the one with the highest second-worst outcome should be chosen. If there is more than one alternative with the highest value of the second-worst outcome, then the third-worst outcomes are compared, etc. [24].

The maximin rule is maximally cautious. The other extreme is represented by the *maximax rule* according to which we should choose one of the alternatives with the highest hope level (level of the best outcome). If the values of the outcomes can be expressed in numbers, then a middle road can be chosen with the help of *Hurwicz's index*. This is a number α between 0 and 1 that expresses the degree of cautiousness (not the degree of pessimism, although that is how it is usually described). The recommendation is to choose an option that maximizes the value of

$$\alpha \times \mathfrak{min}(A) + (1 - \alpha) \times \mathfrak{max}(A)$$

where $\mathfrak{min}(A)$ is the security level and $\mathfrak{max}(A)$ the hope value [18].

Credal set. Given a credal set, we can apply the *maximin expected utility rule* (MMEU). Then for each option we have to find the probability function that gives rise to the lowest expected utility (probability-weighted utility). This is the probabilistic security level of the option in question. The rule requires that we choose an option with the highest possible probabilistic security level [6]. An alternative would be to calculate both the probabilistic security level and the analogous probabilistic hope level, and then apply Hurwicz's index to find a compromise between the two.

Weighted credal set. If we have a weighted credal set, then we can calculate the weighted average of the probabilities, assigning the appropriate weight to each probability. Each weighted average is then multiplied with the corresponding utility, in the usual manner of expected utility calculations. If the weights are second-order

[3]For more information on decision rules, see Chap. 34.

probabilities, then this amounts to using reductive probabilities to calculate expected utilities. The same method can also be used for weights that do not satisfy the probability axioms.

Daniel Ellsberg [4] proposed an adjustment of this rule to make it more cautious. Suppose that we have a measure (such as $\hat{\mathfrak{p}}$) that represents the best probability estimate. We can then weigh it against the probabilistic security level, using an index of the same type that was proposed by Hurwicz. The resulting value can be described as a cautionized variant of expected utility.

Hence, even if we have limited information about probabilities, or none at all, formal representations of uncertainty make it possible to consistently apply decision criteria such as degrees of cautiousness. But of course, when more information is available, it is mostly advisable to apply a decision procedure that makes use of it.

Acknowledgements I would like to thank Richard Bradley and Karin Edvardsson Björnberg for very useful comments on an earlier version of this text.

References and Proposed Readings

Asterisks (∗) indicate recommended readings.

1. Berger, J. O. (1994). An overview of robust Bayesian analysis. *Test, 3*, 5–59.
2. Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological), 30*, 205–247.
3. *Dubois, D., & Prade, H. (2001). Possibility theory, probability theory and multiple-valued logics: a clarification. *Annals of Mathematics and Artificial Intelligence, 32*, 35–66. [Provides a philosophical background to possibility theory and some related approaches.].
4. Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics, 75*, 643–669.
5. Emby, C. (2009). A comparison of elicitation methods for probabilistic multiple hypothesis revision. In V. Arnold (Ed.), *Advances in accounting behavioral research* (Vol. 12, pp. 85–108). Bingley: Emerald Group Publishing.
6. Gärdenfors, P. (1979). Forecasts, decisions and uncertain probabilities. *Erkenntnis, 14*, 159–181.
7. Gärdenfors, P., & Sahlin, N. -E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese, 53*, 361–386.
8. *Gärdenfors, P., & Sahlin, N. -E. (Eds.). (1988). *Decision, probability, and utility: Selected readings*. Cambridge: Cambridge University Press. [Excellent collection with many of the classic texts, in particular on second-order measures.].
9. Ghosh, S., & Mujumdar, P. P. (2009). Climate change impact assessment: Uncertainty modeling with imprecise probability. *Journal of Geophysical Research Atmospheres, 114*, D18113.
10. Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *Journal of Finance, 64*, 549–578.
11. *Hall, J., Twyman, C., & Kay, A. (2005). Influence diagrams for representing uncertainty in climate-related propositions. *Climatic Change, 69*, 343–365. [An introduction to probability intervals with useful examples of how they can be used in practice.]
12. *Halpern, J. (2003). *Reasoning about uncertainty*. Cambridge, MA: MIT Press. [An excellent book-length introduction to the topic of this chapter.]

13. Hansson, S. O. (2008). Do we need second-order probabilities? *Dialectica, 62*, 525–533
14. Hansson, S. O. (2009). Measuring uncertainty. *Studia Logica, 93*, 21–40.
15. Hansson, S. O. (2017). Uncertainty and control. *Diametros, 53*, 50–59.
16. Harsanyi, J. C. (1977). On the rationale of the Bayesian approach: Comments on professor Watkin's paper. In R. E. Butts & J. Hintikka (Eds.). *Foundational problems in the special sciences* (pp. 381–392). Dordrecht: Reidel.
17. Hume, D. ([1739] 1888). *A treatise of human nature*, Reprinted from the original edition in three volumes and edited, with an analytical index, by L.A. Selby-Bigge. Oxford: Clarendon Press.
18. Hurwicz, L. (1951). *The generalized Bayes-minimax principle: A criterion for decision-making under uncertainty* (Cowles Commission Discussion Paper, 355). Chicago. http://cowles.econ.yale.edu/P/ccdp/st/s-0355.pdf.
19. *Kyburg, H. E., Jr. (1992). Getting fancy with probability. *Synthese, 90*, 189–203. [Brief but informative discussion of probability and uncertainty.].
20. Levi, I. (1980). *The enterprise of knowledge*. Cambridge, MA: MIT Press.
21. *Miranda, E. (2008). A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning, 48*, 628–658. [Overview on probability intervals and related approaches.].
22. Moore, R. E. (1979). *Methods and applications of interval analysis*. Philadelphia: SIAM.
23. Savage, L. J. (1972). *The foundations of statistics* (2nd Ed.). New York: Dover.
24. Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.
25. Shafer, G. (1976). *A mathematical theory of evidence.* Princeton: Princeton University Press.
26. Unwin, S. D. (1986). A fuzzy set theoretic foundation for vagueness in uncertainty analysis. *Risk Analysis, 6*, 27–34.
27. *Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning, 24*, 125–148. [A thorough introduction to credal sets and related approaches.].
28. Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.
29. Zadeh, L. (1978). Fuzzy sets as the basis for a theory of possibility. *Fuzzy Sets and Systems, 1*, 3–28.