Chapter 21
Imprecise Probabilities

Seamus Bradley

Abstract This chapter explores the topic of imprecise probabilities (IP) as it relates to model validation. IP is a family of formal methods that aim to provide a better representation of severe uncertainty than is possible with standard probabilistic methods. Among the methods discussed here are using sets of probabilities to represent uncertainty, and using functions that do not satisfy the additvity property. We discuss the basics of IP, some examples of IP in computer simulation contexts, possible interpretations of the IP framework and some conceptual problems for the approach. We conclude with a discussion of IP in the context of model validation.

Keywords Imprecise probabilities · Lower previsions · Credal sets · Formal epistemology · Computer simulation

21.1 Introduction

Model validation is an important aspect of quality control when modelling some phenomenon about which we are uncertain. So, accommodating and representing uncertainty is of central importance to model validation. Probability theory provides the standard suite of tools for dealing with uncertainty, but this theory has its limits. For example, models will often contain parameters whose true value we don't actually know. Now, we can't run a simulation without providing a value for this parameter, so for each simulation we run, we must pick *some* value. If we sample this value from a distribution, and run several simulations—sampling from this distribution each time—we can, to some extent, accommodate uncertainty in the parameter value. In doing so, however, we are assuming that a certain sort of distribution is the "right" distribution to be sampling from. If the parameter fluctuates randomly and

S. Bradley (✉)
TiLPS, Tilburg University, Tilburg, Netherlands
e-mail: s.c.bradley@leeds.ac.uk

S. Bradley
University of Leeds, Leeds, UK

C. Beisbart and N. J. Saam (eds.), *Computer Simulation Validation*,
Simulation Foundations, Methods and Applications,
https://doi.org/10.1007/978-3-319-70766-2_21

we have data on the distribution of fluctuations, perhaps a particular distribution can be justified. If not, it is typical to pick some "non-committal" distribution that will not skew the results of the simulation. For example, if bounds can be put on the range of values the parameter can take, a uniform distribution is often selected. Now, a uniform distribution seems innocuous, non-committal, but the distribution's being uniform for some parameter means that distributions for related parameters are not uniform. Ferson and Ginzburg (1996) give the example of two independent uniformly distributed parameters that give rise to a non-uniformly distributed product. Or consider two inversely related parameters (like "ice fall rate in clouds" and "ice residence time in clouds"): if one is uniformly distributed, the other is not. This is, in essence, the problem that Joseph Bertrand pointed out at the end of the nineteenth century that is today known as "Bertrand's paradox."

The practice of sampling unknown parameters from distributions chosen for convenience rather than for empirically grounded reasons is a necessary aspect of standard modelling practice. Imprecise Probabilities is an approach that attempts to mitigate some of the problematic consequences of such a methodology. This chapter will outline the basic idea of IP, give some examples of IP in modelling contexts, discuss how we might interpret the IP framework and point to some potential problems for IP. We will conclude with a discussion of IP in the context of validation.

21.2 Basics

The core idea of Imprecise Probabilities (IP) is to represent uncertainty using a *set* of probability measures rather than a single such measure (although there are a great many related formal models that we'll discuss later in this section).[1] The basics of uncertainty quantification is introduced in Chap. 5 by Roy and Chap. 22 by Dougherty et al., and the probabilistic/Bayesian approach to uncertainty are discussed in Chap. 7 by Beisbart and Chap. 20 by Jiang et al. in this volume, so let's jump straight to the basic idea of IP. We use **pr** to signify a probability function. The basic idea of IP is that we represent uncertainty, not by a single such function, but by a set of them—$\mathscr{P}$ – defined over the same state space. If X is an event over which the **pr**s are defined, then we can let $\mathscr{P}(X) = \{\mathbf{pr}(X), \mathbf{pr} \in \mathscr{P}\}$. That is, we can take $\mathscr{P}(-)$ to be a set-valued function that returns the set of values assigned to X by members of $\mathscr{P}$. We can then apply the rest of the Bayesian machinery "pointwise." So conditionalising $\mathscr{P}$ involves conditionalising on each member of $\mathscr{P}$ and taking the resultant set of

[1] Although one can find precedents going back to Keynes or even Boole, IP really started in the middle of the twentieth century with work by people like Koopman (1940), Good (1952, 1962), Smith (1961) and Dempster (1967). Work in philosophy on IP really starts with Levi (1974, 1980, 1986). Important formal and philosophical work on IP was carried out by Seidenfeld (1983, 1988); Seidenfeld et al. (1989) (Seidenfeld was Levi's graduate student). Walley (1991) was a hugely influential book which, until recently, was still the go-to monograph for many formal details of the theory. The state of the art in terms of formal theory can be found in Augustin et al. (2014) and Troffaes and de Cooman (2014). Bradley (2014) provides a philosophical overview.

conditional probabilities. Expected values also become sets of values, determined pointwise for each $\mathbf{pr} \in \mathscr{P}$.

Recall that a probability function $\mathbf{pr}$ is a real-valued function on an algebra of events that has the following properties:

Bounded For all X, $0 = \mathbf{pr}(\bot) \leq \mathbf{pr}(X) \leq \mathbf{pr}(\top) = 1$ (where $\bot$ and $\top$ are the bottom and top elements of the algebra, respectively)
Superadditive If $X \wedge Y = \bot$ then $\mathbf{pr}(X \vee Y) \geq \mathbf{pr}(X) + \mathbf{pr}(Y)$
Subadditive If $X \wedge Y = \bot$ then $\mathbf{pr}(X \vee Y) \leq \mathbf{pr}(X) + \mathbf{pr}(Y)$

A function that satisfies the first and second of these properties is called a "lower probability" and a function that satisfies the first and third is called an "upper probability."

The idea of a set of functions and the idea of a lower probability are intimately related. Take a set of probabilities, $\mathscr{P}$ and define $\underline{\mathscr{P}}(X) = \inf \mathscr{P}(X)$, the lowest value assigned to X by some member of $\mathscr{P}$. This function $\underline{\mathscr{P}}$ is a lower probability.[2] Likewise, $\overline{\mathscr{P}}(X) = \sup \mathscr{P}(X)$ is an upper probability. Moreover, $\underline{\mathscr{P}}(\neg X) = 1 - \overline{\mathscr{P}}(X)$.

And going the other way, take a lower probability $\mathbf{lpr}$, and define the associated credal set of $\mathbf{lpr}$ as the set of probability functions such that $\mathbf{pr}(X) \geq \mathbf{lpr}(X)$ for all X, the set of probability functions that "pointwise dominate" it. If $\mathbf{lpr}$ is a lower probability as defined above, such a set is non-empty; let $M(\mathbf{lpr})$ be the associated credal set of $\mathbf{lpr}$. Since $M(\mathbf{lpr})$ is a set of probabilities, we can take the "lower envelope" as we did above: $\underline{M(\mathbf{lpr})}(X)$. The lower envelope theorem entails that $\mathbf{lpr}$ is a lower probability if and only if $\underline{M(\mathbf{lpr})}(X) = \mathbf{lpr}(X)$ for all X (see Sect. 3.3 of Walley 1991 or Sect. 2.2.2 of Augustin et al. 2014). Note that distinct credal sets might result in the same lower probability.[3]

As well as credal sets and lower probabilities, there is a huge range of other related formal methods for representing uncertainty (see, for example, Halpern 2003; Augustin et al. 2014; Klir and Smith 2001). For example, Dempster–Shafer theory (sometimes called Evidence Theory) uses a *belief function* which is a lower probability with the further property of being infinite monotone (a sort of strengthening of superadditivity). DS theory comes equipped with a slightly different interpretation (see Sect. 21.4) and an alternative kind of updating/aggregation rule.[4]

This brief discussion merely scratches the surface of the rich and interesting theory of IP. Many aspects of the statistical method have been replicated inside the IP framework including statistical inference, graphical models (e.g. Bayes nets) and

---

[2]Consider some $\mathbf{pr} \in \mathscr{P}$ for which $\mathbf{pr}(X \vee Y) = \underline{\mathscr{P}}(X \vee Y)$. $\inf \mathscr{P}(X) + \inf \mathscr{P}(Y) \leq \mathbf{pr}(X) + \mathbf{pr}(Y)$, since $\mathbf{pr} \in \mathscr{P}$, so $\underline{\mathscr{P}}$ is superadditive. Boundedness is trivial, and much the same reasoning works if the set $\mathscr{P}$ doesn't attain its bounds (just think in terms of the closure of the set).

[3]There is a one-to-one correspondence between lower probabilities and a subset of the set of credal sets, namely those with some nice topological properties. We don't need to discuss this here, but see the above-listed references for details.

[4]See Oberkampf and Helton (2004) for a discussion of DS theory in an engineering context.

stochastic models (e.g. Markov chains).[5] See, for example Augustin et al. (2014), Troffaes and de Cooman (2014).

21.3 Examples

In this section, we'll explore two examples of IP-like ideas that show up when attempting to model physical systems on computers.

21.3.1 Unknown Parameters

Oberkampf and Roy (2010) discuss an example of how IP arises in scientific computing.[6] They start from the position of wanting to keep apart *epistemic uncertainty* – uncertainty arising from things unknown to the experimenter—and *aleatory uncertainty*—randomness or natural variability.[7] Now, whether some kind of uncertainty counts as one or the other of these kinds is somewhat a matter of perspective, the distinction is important. Aleatory uncertainty about a parameter can be accommodated by having a probability distribution over that parameter in the model. Epistemic uncertainty, on the other hand, is captured by having a set of such distributions, i.e. having a credal set.

The idea is that if a parameter is subject to aleatory uncertainty, you can sample values for that parameter (using the given distribution), run the simulation using those sampled values and then take the distribution of outcome values as telling you something about uncertainty in the outcomes. However, with epistemic uncertainty, you must pick specific values of the parameter to run through the model (and you must pick them using *some* distribution). You can't take the distribution of outputs as telling you about the uncertainty in the outcomes: you can only take the *range* of outcome values as telling you what ranges of values of outcome values are possible given your uncertainty about the parameter. Or perhaps, a more careful way to phrase the same thing: the distribution of output values might be in part due to the choice of input distribution for the unknown paramaters. If that distribution were chosen merely for convenience, then we had better not read too much into the output distribution. As Oberkampf and Helton (2004) say:

> If extreme system responses correspond to extreme values of these parameters (i.e. values near the ends of the uniform distribution), then their probabilistic combination could predict

---

[5]For introductions to these aspects of IP, see Augustin et al. (2014) Chapters 7, 9 and 11, respectively.

[6]I am drawing mainly from Sect.13.4, but similar ideas appear in a number of other places in the book.

[7]This is one dimension of the many ways one can categorise different kinds of uncertainty. See Chap. 5 by Roy in this volume or Morgan and Henrion (1990), Chap. 4.

> a very low probability for such extreme responses. Given that the parameters are only known to occur within the intervals, however, this conclusion is grossly inappropriate (p. 10–3).

As we discussed earlier: the product of independent uniformly distributed variables will not be uniform: it will give more weight to those values in the "middle" of the interval of possible values. If we have no reason to think the variables really are uniformly distributed, then it seems unwise to discount these possible extreme responses as the standard approach implicitly seems to.

This is sometimes known as "probability bounds analysis" or "p-boxes". The following two quotations give you the flavour of this approach.

> In a probability bounds analysis, the uncertainty about the probability distribution for each input variable is expressed in terms of interval bounds on the cumulative distribution function. These bounds form a p-box for each input variable Ferson and Hajagos (2004, p. 136).

> Basically, interval analysis should be used to propagate ignorance, and probability theory should be used to propagate variability Ferson and Ginzburg (1996, p. 133).

A similar approach is advocated by Stainforth et al. (2007b), where they suggest that we interpret the range of values produced by ensemble members to be a "non-discountable envelope" of values of that variable: a range of values that we cannot dismiss as impossible.

21.3.2 The Challenge Problems

In 2002, a workshop was organised around the idea of a set of "challenge problems" that were intended to serve as a kind of standard suite of tests for a theory of uncertainty (Helton and Oberkampf 2004). Oberkampf et al. (2004) presents the challenge problems, and many of the papers in that special issue of the journal respond to them. The problems are designed to highlight issues of "representation, aggregation, and propagation of uncertainty through mathematical models" (Oberkampf et al. 2004, p. 15). The challenge is to come up with some way to predict the behaviour of a system given a model of the system and some evidence as regards some unknown parameters of that system. Each problem has two unknown parameters, and some sort of mathematical model whose output depends on those unknowns. The information about the unknown parameters might be given in a number of different ways. A simple example is in problem 1, we are told that parameter a is somewhere in the interval $[a_1, a_2]$. A more complex example is given by problem 3c where you are told that you have n independent sources of information regarding parameter b, each witness j tells you that b lies in an interval $[b_1^j, b_2^j]$. The model whose outputs depends on the parameters can also be more or less complex. For example, for some of the models, it is simply a function of the parameters. In other cases, the parameters are meant to represent physical constants of some simple physical system.

Several papers have presented broadly IP solutions to this problem set. For example, de Cooman and Troffaes (2004) use the theory of lower previsons to address the problems, while Ferson and Hajagos (2004) use p-boxes.[8]

21.3.3 Nonprobabilistic Odds

Frigg et al. (2014) offer a cautionary tale that suggests that treating distributions of model output as capturing decision-relevant probabilities is dangerous when the target system appears to behave chaotically. The starting point is the scepticism about ensemble forecast probability distributions expressed by Stainforth et al. (2007a):

> The frequency distributions across the ensemble of models may be valuable information for model development, but there is no reason to expect these distributions to relate to the probability of real-world behaviour. One might (or might not) argue for such a relation if the models were empirically adequate, but given nonlinear models with large systematic errors under current conditions, no connection has been even remotely established for relating the distribution of model states under altered conditions to decision-relevant probability distributions (p. 2154).

Frigg et al. (2014) develop a simple example that illustrates this point. They start with a simple mathematical system that an agent wants to use to model a variable of interest. The target system's dynamics are similar to but not identical to the system used for prediction (the one timestep error is always less than one in a thousand). Unfortunately, both the target system and the model exhibit chaotic behaviour, which means that these errors compound and grow. If we are predicting at about eight timesteps out, the errors can grow to such an extent that the distribution of model outputs for an ensemble of nearby initial conditions can be located entirely on the left-hand side of the unit interval while the ensemble of outputs for the target system is entirely on the right-hand side. What this means is that the model appears to be telling you that it's overwhelmingly likely that the variable will be less than 0.5, while the truth is that it's overwhelmingly likely to be greater than 0.5. Obviously betting using these ensemble probabilities would be disastrous. What Frigg et al. (2014) show is that, in fact, it's very often disastrous to bet using ensemble probabilities in a case like this where the dynamics are nonlinear and there's a chance of model error.

They suggest that instead of taking the ensemble probabilities at face value, they should be manipulated to produce "nonprobabilistic odds" which don't yield ruinous betting strategies. How exactly this process should be effected is still up for debate, but what is clear is that the nonprobabilistic odds thus produced will be inversely proportional to upper probabilities, in the same way that probabilistic odds are inversely related to standard probabilities.

[8]See also Fetz and Oberguggenberger (2004) and Helton et al. (2004) for further examples of IP approaches to the challenge problems. See Ferson et al. (2004) for an overview of the range of responses to the challenge problems.

21.4 Interpretations

What does it mean to say that our uncertainty is captured by a set of probability measures (or an upper probability, or a p-box, or…)? In this section, we shall discuss some ways of interpreting such claims. I will discuss several such ideas, but I do not mean to suggest that this survey is exhaustive, nor that the ideas presented here are mutually exclusive: it is certainly possible to be motivated by more than one of these interpretations of the formalism.[9]

21.4.1 One-Sided Betting

Consider betting on the value of a random variable.[10] Let $\mathscr{X}$ be the set of values that random variable X can take. A bet on the value of X can be described by a function from $\mathscr{X}$ to the real numbers. Call such functions *gambles*. How much would you be willing to pay for a gamble g? That is, when do you find the gamble $g - \mu$ desirable? (Where μ is the constant gamble that corresponds to the amount you pay to take g). It seems like the highest price you are willing to pay to take g reflects your valuation of g. Let $\mathbf{lpr}(g) = \sup\{\mu \in \mathbb{R} : g - \mu \text{ is desirable}\}$. Now consider the minimum price you would accept to sell the gamble g. That is, the minimum price at which you find $\mu - g$ desirable. Given some reasonable coherence constraints on your set of desirable gambles,[11] if you require that this should be equal to **lpr**, then **lpr** is a *linear prevision*. And indeed, if we consider gambles over a set of indicator functions for some set of states, then **lpr** gives a probability function. The exploration of linear previsions as a foundation for probability theory goes back to Bruno de Finetti. This is one version of what is known as the "Dutch book theorem," since the coherence constraints on desirability essentially prevent you accepting a collection of bets that guarantee you a sure loss (see Chap. 7 by Beisbart in this volume).

If you drop the requirement that your maximum buying price and your minimum selling price should be the same—if you move away from "two-sided" betting to "one-sided" betting—then **lpr** behaves somewhat like a "lower expectation" operator (called a lower prevision) and its restriction to gambles on indicator functions is a lower probability as defined above. The theory of lower previsions was first systematically set out in Walley (1991), Troffaes and de Cooman (2014) provides an admirably clear self-contained treatment of the theory, as well as significant refinements. Note that, the bets discussed in Frigg et al. (2014) are one-sided bets in this sense.

---

[9]For more on the interpretation of IP, see Bradley (2014).

[10]We earlier described probability theory in terms of events rather than random variables, but the difference is mostly cosmetic. Real-valued random variables are functions from events to real numbers, events are "indicator functions" in the space of random variables.

[11]For example, if you find f desirable, and you find f' desirable, you should find $f + f'$ desirable; or if a gamble's payout is always non-negative, then it is desirable.

If we consider real-world instances of bookmakers or financial traders, there is typically a difference between their buying and selling prices for their commodities (bets, financial products, whatever). Now, part of this spread is explained by the desire to make a profit, but there is evidence that the "bid-ask spread" can also be responsive to the amount of uncertainty about the future performance of the instrument (Smith and van Boening 2008).

So, if we interpret lower previsions and lower probabilities as reflecting the agent's limiting willingness to bet—as is standard in Bayesian approaches—this gives us a natural interpretation of the formalism that is broadly in line with the standard precise probabilist picture.

21.4.2 Indeterminate Belief

One way to interpret credal sets is to take them to reflect an *indeterminacy* in rational belief. If $\mathscr{P}(X)$ is a set of values, this means that it is indeterminate—vague—what rational belief you ought to adopt in X given the evidence that determined that $\mathscr{P}$. This approach takes inspiration from the *supervaluationist* theory of logic, which uses a *set* of truth valuation functions to characterise the satisfaction of a vague predicate. If Wayne is a borderline case of the predicate "bald," then "Wayne is bald" is true according to some members of the set of valuations—"true on some precisifications"—and "Wayne is bald" is false on others (Williamson 1994). Rinard (2013, 2015) has argued for a supervaluationist understanding of credal sets: if it is indeterminate whether the agent's credence in X is stronger than 0.5, some members of the credal set have $\mathbf{pr}(X) > 0.5$ while others $\mathbf{pr}'(X) < 0.5$. If every member of the credal set agrees on something (e.g. that $\mathbf{pr}(X) > 0.1$), then it is determinately true that the agent believes that $\mathbf{pr}(X) > 0.1$. This idea is sometimes characterised by the metaphor of the "credal committee" (Joyce 2010): each probability in the credal set is a committee member and the committee as a whole must decide what to do. When there is unanimity in the committee then things are easy, when there is conflict—disagreement—then things are tricky.[12]

If you want to treat your credal sets or your lower probabilities not as subjective credences but as something akin to objective chances, then you might still be able to take a view of this form: the imprecision in your probabilities is due to objective indeterminacy in the world. This is an underexplored possibility, but see Bradley (2016).

[12]The idea of IP as reflecting unresolved conflict—either between persons or within a person—is one that Isaac Levi discussed in great detail Levi (1980, 1986).

21.4.3 Robustness Analysis

Let's say you run your model with a particular set of parameters, but you are not confident that the parameters you chose are the actual ones. If there's a danger that the result you obtain depends in a big way on the specific value of the parameter chosen, then perhaps, it's best to explore how robust your result is when changing those parameter values. The range of output values, the range captured by the set of probability functions, reflects a robust range of possible outcomes. This "robust bayesian analysis" has a rich history. See, for example Ruggeri et al. (2005). This idea is very closely connected to the discussion of "probability bounds analysis" and "non-discountable envelopes" discussed in Sect. 21.3.1.

If one is taking a Bayesian approach to validation (cf. Chap. 7 by Beisbart and Chap. 20 by Jiang et al. in this volume), then one has to respond to the "problem of the priors": the criticism that Bayesian methods rely on epistemically unmotivated prior probability. One response to such a criticism would be to move to an "imprecise Bayesian" perspective which is, essentially, to apply the robustness analysis approach to the prior. The set of priors allows one to be confident that one's conclusions are not artefacts of the particular prior one chose.

21.4.4 Evidence Theory

Instead of interpreting **lpr** as a degree of confidence, or a limiting willingness to bet, one might want to interpret $\mathbf{lpr}(X)$ as "the degree to which the evidence supports X". This is the interpretation typically associated with the "Dempster–Shafer function" approach to evidence. A "Dempster–Shafer belief function" is a lower probability that has the additional property of being "infinite monotone." The actual formal description of this property is a little messy, and not particularly illuminating in the current context, but see Halpern (2003, Chap. 2.4) for the basics of Dempster–Shafer theory, and see Augustin et al. (2014, Chaps. 4 and 5) and Troffaes and de Cooman (2014, Chaps. 6 and 7) for belief functions and their relation to lower probabilities. Dempster–Shafer theory also has a distinct theory of evidence combination which is beyond the scope of this chapter (but see Halpern (2003, Chap. 3.4)).[13]

The motivating idea behind this degree of support idea is that your evidence can support X to degree p without thereby supporting $\neg X$ to degree $1 - p$ (as would be required if degree of support were probabilistic). Hawthorne (2005) argues that Bayesians need to keep degree of belief and degree of support distinct (and that both concepts are useful).

[13] See also Oberkampf and Helton (2004) for an example of DS theory in an engineering context.

21.5 Problems

IP suffers from a number of issues. Here, we shall outline some of them. As we'll see, not all of them are really that worrying in the context of model validation.

21.5.1 Updating

Recall that earlier, we said that we conditionalise a set of probabilities pointwise. That is, $\mathscr{P}(X|Y) = \{\mathbf{pr}(X|Y), \mathbf{pr} \in \mathscr{P}, \mathbf{pr}(Y) > 0\}$.[14] There are two problems with conditionalising this way that is worth mentioning briefly.

First, dilation. An early, important discussion of dilation is Seidenfeld and Wasserman (1993).[15] Consider a set of probabilities $\mathscr{P}$ constrained as follows for all $\mathbf{pr} \in \mathscr{P}$:

- $\mathbf{pr}(H) = \frac{1}{2}$
- $\mathbf{pr}(H|X) = \mathbf{pr}(H)$

Now consider the proposition Y which is equivalent to $(H \wedge X) \vee (\neg H \wedge \neg X)$. It's easy to show that for all $\mathbf{pr} \in \mathscr{P}$, $\mathbf{pr}(Y) = \frac{1}{2}$. Let $\mathscr{P}$ contain all probability functions over these propositions other than those ruled out by the above constraints. So $\mathscr{P}(X) = [0, 1]$. Note that, it follows from the definition of Y and some basic probability theory that for all $\mathbf{pr} \in \mathscr{P}$, $\mathbf{pr}(H|Y) = \mathbf{pr}(X)$. Therefore $\mathscr{P}(H|Y) = [0, 1]$. Note that the same reasoning entails that $\mathscr{P}(H|\neg Y) = [0, 1]$. This, in essence, is the phenomenon of dilation. What is this considered a problem? To see this, let's consider an example that gives some meaning to the variables.[16]

I have two coins, one fair and one mystery coin of unknown bias. Let H be the event that the fair coin lands heads, and let X be the event that the mystery coin lands heads. (Verify that the above discussed probabilistic constraints seem reasonable given this interpretation of the propositions). Now I toss both coins and announce that the two coins landed the same way up (either both heads or both tails), call this proposition Y. What is your posterior in the fair coin having landed heads? $\mathscr{P}(H|Y) = [0, 1]$. Your belief in the fair coin's having landed heads has *dilated*: the interval of probability values has spread out from $\{\frac{1}{2}\}$ to $[0, 1]$. And this happens regardless of whether you learn Y or $\neg Y$. This seems puzzling. Learning the fact that the two coins landed the same way up doesn't seem like it should cause me to change my belief in whether the fair coin landed heads. It seems like you have learned something irrelevant to H and it has caused you to become more uncertain about H. That seems like a strange way to arrange your credences.

[14] The restriction to non-zero probability in the conditioning event is for convenience: if we had defined credal sets in terms of Popper functions or similar we could do without such a restriction.

[15] A recent characterisation of dilation is found in Pedersen and Wheeler (2014).

[16] This description of the puzzle follows Joyce (2010).

Dilation, despite initial appearances, isn't as problematic as some (e.g. White (2010)) take it to be. For example, in different ways, Joyce (2010), Bradley and Steele (2014b), Pedersen and Wheeler (2014) and Hart and Titelbaum (2015) all argue that dilation is actually the correct response to the evidence as specified. Gong and Meng (2017) argue that dilation is a symptom of a mis-specified statistical inference problem, not a problem for IP *per se*.

We can think of the problem as follows. The constraints we placed on our model place no constraint at all on what values $\mathbf{pr}(H|Y)$ might take. It is somewhat intuitive that H and Y say nothing about each other. If we take this "silence" to be modelled by probabilistic independence—$\mathbf{pr}(H|Y) = \mathbf{pr}(H) = \frac{1}{2}$ – then our $\mathscr{P}$ becomes the singleton with $\mathbf{pr}(X) = \mathbf{pr}(H|Y) = \frac{1}{2}$. But, as Pedersen and Wheeler (2014) point out, independence for sets of probabilities can be much more subtle (Cozman 2012). As Bradley (2014) explains, probabilistic independence is not the appropriate characterisation of the "silence" of Y with respect to H. As it stands, it is compatible with the problem set up that Y would be very informative about H, if only we knew something about the bias of the mystery coin. That is, if we knew that the mystery coin was biased towards heads, learning that the coins landed the same way up would be evidence in favour of heads on the fair coin. It is not that H and Y are unrelated: it's just that the nature of their relationship is unknown.

Let's turn now to another puzzle related to updating sets of probabilities: the problem of *belief inertia*. This problem, though not under that name, goes back to Sect.13.2 of Levi (1980), and is also discussed by Walley (1991); Vallinder (2018) provides a nice discussion of the current state of the art. Consider the mystery coin again. Recall that proposition X is "mystery coin lands heads up", and $\mathscr{P}(X) = [0, 1]$. Now consider learning that in ten flips of the mystery coin, 8 were heads. Call this proposition Z. This seems like some evidence that could potentially move your credences about. But note that $\mathscr{P}(X|Z) = (0, 1)$. Why? Because, even if we assume that all the priors in $\mathscr{P}$ are "well behaved" beta distributions over the unknown bias of the mystery coin, there are some distributions in $\mathscr{P}$ that put so much weight on the probability for landing heads being really really low that even evidence Z doesn't move them very far away from 0. In the case that $\mathscr{P}(X) = [0, 1]$ the "credal committee" contains members that are so stubborn that they are moved an arbitrarily small distance by the evidence. And likewise for the top end of the unit interval. Starting with a vacuous prior like this seems to make learning impossible.

The imprecise probabilities that are likely to arise in a validation setting are not likely to be vacuous, so perhaps this is less of a concern in the current context.

21.5.2 Decision-Making

Ultimately, we often want to use the results of our simulations for decision support: we want to take our simulation of the behaviour of a nuclear reactor to inform safety standards for new reactors, for example. This boils down to the question: how do we translate our uncertain predictions into policy advice? We want to take into account the uncertainty in our simulations and perhaps err on the side of caution by focusing on, for example the worst case among the plausible scenarios consistent with our evidence. So how do we make decisions with sets of probabilities?

If you had a single probability function, you can act so as to maximise expected utility. What is the analogue decision rule for imprecise probabilities? There are a number of possibilities, each with drawbacks. Should you act to maximise minimum expectation over probabilities in your set? Pick an option that maximises expected value with respect to some $\mathbf{pr} \in \mathscr{P}$?[17] Find some way to average over the set of expectation and maximise that?[18] Elga (2010) argues that no imprecise decision theory is even minimally adequate, although the consensus now seems to be that Elga overstated his case (Bradley and Steele 2014a; Chandler 2014; Sahlin and Weirich 2014). In any event, it is still true that providing IP with an adequate decision theory is an unresolved issue. In a sense, it is not surprising that decision-making with IP is difficult: the whole point is that we are being careful to represent the full extent of our lack of knowledge, and we shouldn't expect decision- making to be easy in such a case. Indeed, it would appear to be a surprise if decision-making were as easy as it is in a case where we know the objective probabilities of the events, or have some reason to believe our subjective estimates are on the right track.

21.6 Validation and IP

So where does all of the above leave the practitioner? What should someone who works with computer simulations take away from this discussion of imprecise probabilities? We'll discuss the issues of interpretation and the problems in the sections below, but first, I want to say something about where to situate IP in general. Some, particularly in philosophy, seem to see IP as a rival to the standard Bayesian view of subjective probability. I think this is a mistake. IP is a suite of tools, a range of methods that extend and improve on the standard probabilistic tools. They are provided in order to overcome some problems that the standard theory has with severe uncertainty, careful propagation of uncertainty and giving appropriate weight to serious dangers. Questions remain about when it is appropriate to deploy the admittedly more complex machinery of IP, and when it is best to stick with the simpler tools of standard probability, but the above discussion of the "challenge problems" highlights that many practictioners do see value in the use of IP.

[17] This option, called "E-admissibility" by Levi (1974)—and discussed in depth in Levi (1986)—is a popular one among some IP theorists.

[18] See Bradley (2015) for some discussion of the options.

21.6.1 Interpretations

We discussed four ways of interpreting the mathematical framework of IP: betting, indeterminate belief, robustness analysis and evidence theory. Which of these will be appropriate depends on your goals and how you are using the tools. Probabilities show up in validation, and there is a question about how to interpret them.[19] Your general interpretative view on probabilities is going to inform what you think about IP. If your inclination is to treat probabilities in a subjective Bayesian way, then the betting approach seems a natural fit for interpreting imprecise probabilities. If, however, your inclination is to treat probabilities as objective chances or objective evidential probabilities, then perhaps the "indeterminate belief/chance" route is a better fit.

If the imprecise probabilities arise from responding to the aleatory/epistemic uncertainty distinction in an "unknown parameters" context (see Sect. 21.3.1) then it makes sense to see IP as a kind of "robustness analysis".

The case of the "evidence theory" view of IP—$\mathbf{lpr}(X)$ represents the degree to which the evidence supports X—is an interesting one. This interpretation is strongly associated with Dempster–Shafer belief functions which are a special case of lower probabilities. It seems that much of the literature on this topic doesn't make a distinction between the "robustness analysis" view and the "evidence theory" view.

21.6.2 Problems

We looked at two kinds of problems for IP: problems related to updating and problems related to decision- making.

It seems that the problems for updating aren't all that problematic in the validation applications of IP. First, note that conditionalisation plays a relatively minor role outside of the Bayesian approach. And even where conditionalisation does play a role, the goal is to propagate the uncertainty. So if that yields large intervals of output variables then that is a good thing: the uncertainty has been adequately propagated. Take dilation: dilation occurs when you have two variables whose interaction is unknown. By that I mean, when it is unknown whether they are positively or negatively correlated.[20] In a validation context, if you were in that situation, you would want to know the range of possible system responses if the parameters were positively correlated or if they were negatively correlated: you would *want* to see that range of responses represented in your model output. Likewise for belief inertia. If it is consistent with your evidence that X might not be affected very much by conditioning on Z, then your model output should accommodate that possibility. In short, the first two problems discussed are problematic when we interpret $\mathscr{P}(A|B)$ as rational credence in A having learned B, but in a context where the probability models are representing

[19]See Hájek (2011) for an introduction to interpretations of probability.

[20]For a more careful and rigorous characterisation of dilation, see Pedersen and Wheeler (2014).

possible relationships between variables in a model, the phenomena of dilation and inertia do not seem so problematic.

So what about decision-making? If we are required to make a decision on the basis of some simulation-based prediction that involves several underconstrained parameters, it would be a mistake to make decision-making too easy. Take a simple example of deciding on how high to build your flood defences. Let's imagine that you use a climate model to predict whether sea level rises will be small, moderate or large. We run a bunch of climate models, varying some unknown parameters, and come up with a range of possible future scenarios. A precise probabilistic approach might say that on average sea level rise will be moderate. Now, it might be tempting to build flood defences that can cope with moderate rise but not with a large rise. It might also be tempting to not bother with investing in defences that can be extended in the case of a large rise: after all, the models say that won't happen. Of course, such a decision maker has made a mistake in paying too much attention to the headline "moderate rise" and not enough attention to the small print "range of scenarios". By taking seriously the task of propagating the uncertainty and presenting the full range of scenarios consistent with the physical constraints, the IP approach highlights the ranges of cases that the decision maker cannot discount in her deliberations. That makes it harder to make a decision, but it does so in a way that will improve the decisions made. Of course, one can go too far: the range of climate scenarios consistent with our models range from ice age to everybody dies of heatstroke. It's hard to make any sort of decision that will lead to good outcomes across the board there. But propagating that uncertainty, presenting the "non-discountable envelope" of scenarios, prevents the decision maker from being misled by the headline model average.

21.7 Conclusion

Imprecise probabilities can provide an expressively rich and sophisticated theory of uncertainty that builds on and extends orthodox probability theory. The formal foundations of IP are fairly solid although there are still some conceptual sticking points that need work. Such a theory has the potential to find many useful applications in the field of computer simulation validation.

References

Augustin, T., Coolen, F. P., de Cooman, G., & Troffaes, M. C. (Eds.) (2014). *Introduction to imprecise probabilities*. John Wiley and Sons.

Bradley, S. (2014). Imprecise probabilities. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy*.

Bradley, S. (2015). How to choose among choice functions. In *ISIPTA 2015 proceedings* (pp. 57–66).

Bradley, S. (2016). Vague chance. *Ergo*, *3*(20)

Bradley, S., & Steele, K. (2014a). Should subjective probabilities be sharp? *Episteme*, *11*, 277–289.

Bradley, S., & Steele, K. (2014b). Uncertainty, learning and the "problem" of dilation. *Erkenntnis*, *79*, 1287–1303.

Chandler, J. (2014). Subjective probabilities need not be sharp. *Erkenntnis*, *79*, 1273–1286.

Cozman, F. (2012). Sets of probability distributions, independence and convexity. *Synthese*, *186*, 577–600.

de Cooman, G., & Troffaes, M. (2004). Coherent lower previsions in systems modelling: Products and aggregation rules. *Reliability Engineering and System Safety*, *85*, 113–134.

Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, *38*, 325–339.

Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers' Imprint*, 10.

Ferson, S., & Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety*, *54*, 133–144.

Ferson, S., & Hajagos, J. G. (2004). Arithmetic with uncertain numbers: Rigorous and (often) best possible answers. *Reliability Engineering and System Safety*, *85*, 135–152.

Ferson, S., Joslyn, C. A., Helton, J. C., Oberkampf, W. L., & Sentz, K. (2004). Summary from the epistemic uncertainty workshop: Consensus amid diversity. *Reliability Engineering and System Safety*, *85*, 355–369.

Fetz, T., & Oberguggenberger, M. (2004). Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliability Engineering and System Safety*, *85*, 73–87.

Frigg, R., Bradley, S., Du, H., & Smith, L. A. (2014). Laplace's demon and the adventures of his apprentices. *Philosophy of Science*, *81*, 31–59.

Gong, R., & Meng, X. -L. (2017). Judicious judgment meets unsettling update: Dilation, sure loss and simpson's paradox. arXiv:1712.08946v1.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society Series B (Methodological)*, *14*, 107–114.

Good, I. J. (1962). Subjective probability as the measure of a non-measurable set. In *Logic, methodology and philosophy of science: Proceedings of the 1960 international congress* (pp. 319–329).

Hájek, A. (2011). Interpretations of probability. In E. N. Zalta (ed.), *The stanford encyclopedia of philosophy*. Stanford. http://plato.stanford.edu/archives/fall2007/entries/probability-interpret/.

Halpern, J. Y. (2003). *Reasoning about uncertainty*. MIT Press.

Hart, C., & Titelbaum, M. G. (2015). Intuitive dilation? *Thought*, *4*, 252–262.

Hawthorne, J. (2005). Degree-of-belief and degree-of-support: Why Bayesians need both notions. *Mind*, *114*, 277–320.

Helton, J. C., Johnson, J., & Oberkampf, W. L. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering and System Safety*, *85*, 39–71.

Helton, J. C., & Oberkampf, W. L. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety*, *85*, 1–10.

Joyce, J. M. (2010). A defense of imprecise credence in inference and decision. *Philosophical Perspectives*, *24*, 281–323.

Klir, G. J., & Smith, R. M. (2001). On measuring uncertainty and uncertainty-based information: Recent developments. *Annals of Mathematics and Artificial Intelligence*, *32*, 5–33.

Koopman, B. O. (1940). The bases of probability. *Bulletin of the American Mathematical Society*, *46*, 763–774.

Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy*, *71*, 391–418.

Levi, I. (1980). *The enterprise of knowledge*. The MIT Press.

Levi, I. (1986). *Hard Choices: Decision making under unresolved conflict*. Cambridge University Press.

Morgan, M. G. & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press.

Oberkampf, W. L., & Helton, J. C. (2004). Evidence theory for engineering applications. In *Engineering design reliability handbook* (pp. 197–226). CRC Press.
Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., & Ferson, S. (2004). Challenge problems: Uncertainty in system response given uncertain parameters. *Reliability Engineering and System Safety*, *85*, 11–19.
Oberkampf, W. L., & Roy, C. J. (2010). *Verification and validation in scientific computing*. Cambridge University Press.
Pedersen, A. P., & Wheeler, G. (2014). Demystifying dilation. *Erkenntnis*, *79*, 1305–1342.
Rinard, S. (2013). Against radical credal imprecision. *Thought*, *2*, 157–165.
Rinard, S. (2015). A decision theory for imprecise probabilities. *Philosophers' Imprint*, *15*, 1–16.
Ruggeri, F., Insua, D. R., & Martín, J. (2005). Robust bayesian analysis. In *Handbook of statistics* (Vol. 25, pp. 623–667). Elsevier.
Sahlin, N.-E., & Weirich, P. (2014). Unsharp sharpness. *Theoria*, *80*, 100–103.
Seidenfeld, T. (1983). Decisions with indeterminate probabilities. *The Brain and Behavioural Sciences*, *6*, 259–261.
Seidenfeld, T. (1988). Decision theory without "independence" or without "ordering". what's the difference? *Economics and philosophy* (pp. 267–290).
Seidenfeld, T., Kadane, J. B., & Schervish, M. J. (1989). On the shared preferences of two Bayesian decision makers. *The Journal of Philosophy*, *86*, 225–244.
Seidenfeld, T., & Wasserman, L. (1993). Dilation for sets of probabilities. *Annals of Statistics*, *21*, 1139–1154.
Smith, C. A. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society Series B (Methodological)*, *23*, 1–37.
Smith, V. L. & van Boening, M. (2008). Exogenous uncertainty increases the bid-ask spread in the continuous double auction. In *Handbook of experimental economics results* (vol. 1). Elsevier.
Stainforth, D. A., Allen, M. R., Tredger, E., & Smith, L. A. (2007a). Confidence uncertainty and decision-support relevance in climate models. *Philosophical Transactions of the Royal Society*, *365*, 2145–2161.
Stainforth, D. A., Downing, T., Washington, R., Lopez, A., & New, M. (2007b). Issues in the interpretation of climate model ensembles to inform decisions. *Philosophical Transactions of the Royal Society*, *365*, 2163–2177.
Troffaes, M., & de Cooman, G. (2014). *Lower previsions*. Wiley.
Vallinder, A. (2018). Imprecise bayesianism and global belief inertia. *The British Journal for the Philosophy of Science*, *69*, 1205–1230.
Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities, volume 42 of Monographs on Statistics and Applied Probability*. Chapman and Hall.
White, R. (2010). Evidential symmetry and mushy credence. In T. Szabo Gendler, & J. Hawthorne (Eds.), *Oxford studies in epistemology* (pp. 161–186). Oxford University Press.
Williamson, T. (1994). *Vagueness*. Routledge.