**Review Article**

Hykel Hosni* and Jürgen Landes

# Logical perspectives on the foundations of probability

**Abstract:** We illustrate how a variety of logical methods and techniques provide useful, though currently underappreciated, tools in the foundations and applications of reasoning under uncertainty. The field is vast spanning logic, artificial intelligence, statistics, and decision theory. Rather than (hopelessly) attempting a comprehensive survey, we focus on a handful of telling examples. While most of our attention will be devoted to frameworks in which uncertainty is quantified probabilistically, we will also touch upon generalisations of probability measures of uncertainty, which have attracted a significant interest in the past few decades.

**Keywords:** uncertainty, logic, probability, artificial intelligence, events, coherence, induction

**MSC 2020:** 03-01, 03-03, 03B42, 03B48, 60-01, 60A99, 68T27, 68T30, 68T37

## 1 Introduction

The key problem in uncertain reasoning we consider in this article is as follows:

KEY PROBLEM

**Given** A set of events $\Gamma = \{\gamma_1, \ldots, \gamma_k\}$ of interest.

**Want** The quantification of the agent's uncertainty on $\Delta \supseteq \Gamma$ subject to the following constraints:

(1) All the available information is represented adequately.

(2) All the remaining uncertainty is quantified adequately.

It is commonly held in science that probability theory, and its statistical and computational applications provide the mathematical toolkit of choice to tackle the wide range of specific instantiations of the KEY PROBLEM. Such a common view can be seen as the fulfilment of early visions put forward by Laplace and Maxwell, among others. Paraphrasing slightly, the former claimed that probability amounts to commonsense reduced calculus, whereas the latter noted that probability is the only logic scientists really need.

Likening probability to "the logic of science," as the subtitle of [1] puts it, is indeed a commonplace. However, in many context of scientific and practical interest, there is much to be gained by taking logic literally, rather than metaphorically. Hence, in this note, we set ourselves the goal of suggesting that logical methods can provide very useful tools, albeit currently underappreciated, for addressing the aforementioned KEY PROBLEM. As it will be apparent, logic plays a fundamental role in addressing the two constraints.

First, we emphasise the role of logic as a tool of choice for *knowledge representation*, which arises as a consequence of the semantics governing the formal notion of *event*. Taking a logical (rather than the usual set-theoretic) perspective on events makes the representation of imperfect knowledge mathematically natural and conceptually robust. This leads to the pivotal role of logic to capture the second constraint of the KEY PROBLEM.

* **Corresponding author: Hykel Hosni,** Department of Philosophy "Piero Martinetti", University of Milan, Milan, Italy,
e-mail: hykel.hosni@unimi.it

**Jürgen Landes:** Department of Philosophy "Piero Martinetti", University of Milan, Milan, Italy, e-mail: juergen_landes@yahoo.de

The many logical semantics available provide a set of *uncertainty resolution methods*, each focussing on a specific aspect of the representation of the subtle relation between information and uncertainty. The construction of a coherent space of logical possibilities is arguably a necessary condition for successful applications of uncertain reasoning in practical problems. Indeed, as we will argue in the last part of this note, logic with its native focus on *inference* and *computation* can shed very important light on the pressing problem of endowing artificial intelligence (AI) systems with transparent knowledge representation and reasoning (KRR) abilities. As a telling example, we will see how in practice, probabilistic (conditional) independencies are often present in problem specifications. Bayesian networks have become a popular tool in AI exploiting them for inferential and computational gains.

Our review is free of formality and clearly partial, in the sense that it reflects our competence and interests. In writing it we have aimed at providing readers with a flavour of those key roles that logic can play in the wider field of uncertain reasoning. We have made our selection of cited literature wide so as to provide a mix of conceptual, mathematical, and practical relevance. This hopefully provides a sufficiently varied list of key pointers to relevant literature.

## 1.1 Motivating the logical perspective

It is remarkably difficult to explain to what logic is for, as van Benthem put it in [2],

> The unity of logic, like that of other creative disciplines, resides in the mentality of its practitioners and their modus operandi.

And of course, mentality varies a lot, in time, and within each of the many academic silos. To illustrate, consider the logical landscape of the late 1920s, i.e. when Bruno de Finetti begun focussing on the coherentist foundation of probability recalled below in Section 1.3. For him, logic *was* the Boolean algebra of sets, which he interpreted as the logic of certainty. No alternative was available to him. He later became aware of the introduction of many-valued logics by the Polish school, but he assessed it as an alternative to probability, not to Boolean logic [3]. Fifty years on, the emergence of expert systems in AI pressed logicians to take seriously the idea that reasoning may not necessarily be grounded in Boolean logic. From fuzzy logics to dynamic epistemic logics, non-monotonic logics, and paraconsistent logics, a galaxy of extensions of classical logic have been thoroughly investigated towards the end of the past century [4]. Against this background, one in which many logics are worth taking as the starting point of an agent's knowledge representation, logicians could ask whether de Finetti's coherence-based justification for measuring uncertainty with probability necessitated logic to be classical. In a landmark contribution, Paris [5] showed that this was indeed not the case: many logics can yield distinct instantiations of *coherence*, as we detail in Section 2. Accordingly, the logical perspective we put forward in this note is not tied to a specific logical system, but aims at illustrating, by way of example, what the *modus operandi* of logicians is, and how it could be beneficial to the wider field of uncertain reasoning.

Logic and probability concur to achieving the goal of representing both an agent's knowledge and their uncertainty in events of interest, thereby providing a guideline for how to *act rationally* under the specific circumstances. This was anticipated by Leibniz in his *Nouveaux essais* where he writes:

> I maintain that the study of the degrees of probability would be very valuable and is still lacking, and that is a serious shortcoming in our treatises on logic. For when one cannot absolutely settle a question one could still establish the degree of likelihood on the evidence, and so one can judge rationally which side is the most plausible.[1]

The decision-theoretic foundation of probability goes back at least to Huygens's analysis of expectation, see, e.g. [6], and has been put forward explicitly by the so-called Bayesian school [7–11]. They did so by relying essentially on the logical notion of *coherence*, which, however, has been investigated with logical tools only fairly recently and still appears to be underappreciated. Indeed, logic and probability constitute mostly independent elements

---

**1** Translation by P. Remnant e J. Bennet (1981) p. 372. See [12] for more background.

of today's mathematical and scientific education. This is at odds with the fact that they had a joint start as *mathematical theories of reasoning* in the mid-1800s – see [13,14] for a thorough historical reconstruction.

De Morgan, who coined the term *mathematical logic*, titled his volume *Formal Logic: Or the calculus of inference, necessary and probable* [15]. His friend and collaborator, George Boole, gave his *Laws of Thought* a telling subtitle: *On which are founded the mathematical theories of logic and probability*. And indeed the KEY PROBLEM originates in Boole's seminal volume, where it is claimed to be solvable by a "general method":

> the final expression will contain terms with arbitrary constant coefficients [so] by giving to their constants their limiting values 0 and 1 determine the limits within which the probability sought must lie independently of all experience." [16] p. 17.

Focussing on such a general method was a breakthrough in both logic *and* probability. Before him, indeed, logic was mostly concerned with the analysis of individual patterns of reasoning – syllogisms. After him, logicians set sail to algorithmic reasoning. As to probabilities, however, Boole noted that whenever the information we have is "insufficient to render determinant the value sought," the "general method" yields (probability) intervals, rather than values, which have been tightly pinned down by Maurice Fréchet in 1935 as follows. If $\theta, \phi$ are events (see below for notation and terminology) with probabilities $P(\theta) = x$ and $P(\phi) = y$, then the probability of the disjunction and the probability of the conjunction of $\theta$ and $\phi$ are bound by

$$\max\{x, y\} \leq P(\theta \vee \phi) \leq \min\{1, x + y\}; \tag{1}$$

$$\max\{0, x + y - 1\} \leq P(\theta \wedge \phi) \leq \min\{x, y\}, \tag{2}$$

see [17] for an early detailed analysis.

The aforementioned inequalities provide a useful starting point in our attempt to present a vast body of ideas in a unified matter: pinning down the "logic of science" by combining logic with probability raises, from the outset, a number of questions which result in a large space of formal possibilities. Logical methods in our view provide fundamental tools to navigate this space. Firstly, the aforementioned bounds show that, in general, the quantification of an agent's uncertainty may not be uniquely determined by the available information. This ties in with the key questions tackled in (pure) inductive logic, the topic of Section 3. Second, and somewhat alternatively, whenever uncertainty is not uniquely quantified by the available information, it may be methodologically more appropriate to consider the *set* of probability distributions compatible with it, as a justified measure of uncertainty. And, very interestingly, sets, rather than single probability functions, have been shown to emerge naturally by grounding the formalism on extensions of Boolean logics, as pointed out in Section 2.

Before going into details, readers with no logical background may benefit from a short guided tour of logic in the theory of uncertain reasoning, to which the remainder of this section is devoted. In it, we also lay down a bare minimum of terminology and notation.

## 1.2 Basic notions and terminology

### 1.2.1 Events

Analysing *events* is a key contribution of logic to the theory of probability. That such an analysis could be done independently of any underlying probabilistic experiment, forecast or decision, was one of Boole's visionary intuitions. Far from being perfect, his work [16] was nonetheless a convincing defence that propositional or *Boolean* logic, as we now know it, serves as a common basis for both logic and probability. Among those who were not particularly impressed, it is ironic that we can list E. Jaynes who, as we noted earlier, terms probability as "the logic of science":

> [Boole's] work on probability theory contains ludicrous errors, far worse than any committed by Laplace [...]. While Laplace considered real problems and got scientifically useful answers, Boole invented artificial school-room type problems, and often gave absurd answers. [18], p. 242.

It should be noted that most of such "absurdities" are to be found in connection to conditional probability, which is indeed very hard to cast logically, see [19,20] for recent promising results in the field.

In spite of the lack of rigour, and indeed the many mistakes to be found in Boole's seminal book, it turned out that Boolean algebras are the particular formulation of the logic which serves the purpose particularly well, see, e.g. the comprehensive review [21]. In elementary cases, two Boolean algebras are at work: **A** formalising events and **2** formalising their indicator functions. Homomorphisms from **A** to **2**, corresponding to the logical notion of "valuation" (see below), then represent the uncertainty-resolving information available to the agent, so that the probability unit mass can be distributed accordingly. In short, Boolean semantics provides the tools to represent the distinction between what is known and what is not known to an agent about a specific event of interest.

We thus identify events, the bearers of probability, with the elements $\theta$, $\phi$, etc. of the set of *sentences* $\mathcal{SL}$ generated recursively from a (finite) propositional language $\mathcal{L} = \{p_1, ..., p_n\}$, by means of the connectives $\{\neg, \wedge, \vee\}$. As a bit of useful terminology, if $\psi = (\theta \wedge \phi) \in \mathcal{SL}$, then both $\theta$ and $\phi$ belong to $\mathcal{SL}$ and are called *the immediate subsentences* of $\psi$ – similarly for negation and disjunction.

A **Boolean (also known as propositional also known as classical) valuation** is a function $v$ from $\mathcal{SL}$ to $\{0, 1\}$, the set of truth values. Valuations are the key component of logical **semantics**. The characteristic property of Boolean semantics is known as **compositionality**, a property of valuations to the effect that the truth value of a sentence in $\mathcal{SL}$ is a function of the truth value of its immediate sub-sentences, as fixed by the following conditions:

$$
\begin{aligned}
v(\neg\phi) &= 1 - v(\phi), \\
v(\theta \wedge \phi) &= \min\{v(\theta), v(\phi)\}, \\
v(\theta \vee \phi) &= \max\{v(\theta), v(\phi)\}.
\end{aligned}
\tag{3}
$$

We say that $\phi$ is a **(Boolean) logical consequence** of $\theta$, written, $\theta \vDash \phi$ if for all valuations $v$ such that $v(\theta) = 1$, we have $v(\phi) = 1$. In particular, we say that $\phi$ is a **tautology** if $\vDash \phi$, i.e. if $v(\phi) = 1$, for all valuations $v$. The fact that tautologies are true *no matter what*, makes them a strong candidate for the logical formalisation of the **certainly true event**. Compositionality immediately holds that the **certainly false event** is the negation of a tautology.

Note that as a consequence of compositionality, for any $\phi \in \mathcal{SL}$, we have $v(\phi) \in \{0, 1\}$. This corresponds to the (modelling) assumption to the effect that every event *is* either true or false in the model. However, an agent may or may not know, at a particular time, which truth value is actually taken by $\phi$. This temporary ignorance, for which no Boolean formalisation is available, is crucial in understanding the relation between the uncertainty resolution represented by (Boolean) logical semantics, and probability. For two cases are of interest:

(1) $\phi$ is true and the agent knows it. Then the agent will be *certain* that $\phi$ is true. Similarly, if $\phi$ is false and the agent knows it, then it will be *certain* that $\phi$ is false.

(2) The agent does not know whether $\phi$ is true or false. Then the agent is *uncertain* about it.

As a consequence, it is practically convenient and conceptually helpful to call "uncertainty" the state of mind of an agent who is "ignorant" about an event's actual truth value. In this spirit, we say that an agent who is uncertain about an event $\phi$ quantifies their uncertainty about it by making a **forecast**. Hence, a forecast can be thought of as a map $\Phi$ from $\mathcal{SL}$ to $[0, 1]$, where $\Phi(\phi) = 1/0$ is interpreted as the forecast that $\phi$ will/will not happen.

We are of course interested in *rational* forecasts, i.e. forecasts which can guide us successfully in decision making and truth tracking. A variety of results have been put forward to the effect that a forecast meets the demands of rationality if it is compatible with the Kolmogorov axioms for probability. We will go back to this shortly. For the time being, note that forecasts need not be necessarily future oriented. Borel [22] illustrated this by pointing out that it makes perfect sense to be uncertain about the outcome of a coin tossing *after* the coin has actually been tossed, but *before* the relevant observation.

### 1.2.2 Boolean probability functions

Against this background, **an $\mathcal{L}$-probability function** $P$ can be defined as a map $P$ from $S\mathcal{L}$ to the real unit interval such that

– $P$ is *normalised* on contradictions and tautologies, i.e.

$$\text{if } \vDash \phi, \text{ then } P(\phi) = 1; \quad \text{if } \vDash \neg\phi, \text{ then } P(\phi) = 0, \tag{4}$$

– $P$ is *monotone* with respect to Boolean consequence, i.e.

$$\text{if } \theta \vDash \phi \text{ then } P(\theta) \leq P(\phi), \tag{5}$$

– $P$ is *finitely additive*, i.e.

$$\text{if } \vDash \neg(\theta \wedge \phi) \text{ then } P(\theta \vee \phi) = P(\theta) + P(\phi). \tag{6}$$

The aforementioned requirements, which are standard in the field, see e.g. [23], provide the simplest, albeit slightly redundant, logical formulation of the Kolmogorov axiomatisation. Logical valuations which satisfy (3) clearly satisfy (4), (5), and (6), as noted by Kolmogorov himself [24, p. 2]. Note that the converse holds as well when probability values are restricted to the binary set.

The fact that Boolean logic and probability functions can be rooted in a common syntax, in addition to the just noted fact that extreme probability values are Boolean truth values, opens up a rich formal and conceptual interplay between the theories of probability and logic, which is our present focus.

### 1.2.3 Degrees of probability vs degrees of truth

$\mathcal{L}$-probability functions lack, in general, the property of compositionality which Boolean logic enjoys through (3), and which can be generalised well beyond the Boolean case. In contrast to this, it is apparent from equations (2) that the value of $P(\theta \wedge \phi)$ fails to be uniquely determined by those of $P(\theta)$ and $P(\phi)$ (and similarly for disjunctions). This key difference between truth functions and probability functions marks a point of departure between the semantics of logic and probability. This is not undesirable. To see it, suppose probability functions were indeed compositional, in the sense of there being a function $F_\wedge : [0, 1] \times [0, 1] \to [0, 1]$ such that for all events $\theta$ and $\phi$, $P(\theta \wedge \phi) = F_\wedge(P(\theta), P(\phi))$. Now take $P(\theta) = P(\neg\theta) = 1/2$. Then by substitution of equal values, we would obtain

$$P(\theta \wedge \theta) = F_\wedge(P(\theta), P(\theta)) = F_\wedge(P(\theta), P(\neg\theta)) = P(\theta \wedge \neg\theta). \tag{7}$$

By normalisation, we have that $P(\theta \wedge \neg\theta) = 0$ a value which (7) would then force on $P(\theta \wedge \theta)$ which logic forces here to equal $1/2$.

The elementary logical setting of $\mathcal{L}$-probability function here serves the important methodological role of distinguishing two dimensions of uncertainty which can be taken to be largely independent. The former has to do with so-called *degrees of truth*, as they are captured by a number of many-valued logics, to be recalled below. The latter is to do with *degrees of belief*, as they are quantified by probability functions and their non-additive extensions. This distinction, being the logical prerequisite for the coherent combination of the relative formalisms, is particularly important in the field of uncertainty in AI where multiple sources of uncertainty must be processed by the same system, see [4,25]. We will come back to it in Section 2.

## 1.3 Coherence

Armed with $\mathcal{L}$-probability functions, we can ask a set of increasingly more practical questions, all contributing to our opening KEY PROBLEM:

(1) What does it mean to make a *rational* forecast?

(2) Is the answer to the previous question unique?

(3) How can uncertainty representation methods incorporate a variety of kinds of information?

(4) How does uncertain reasoning relate to decision-making?

As it will be apparent, both logical methods and techniques provide sharp analytic tools, and in some cases, the very language in which the questions can be asked in a purposeful way.

So, what, if anything, guarantees that $\mathcal{L}$-probability functions provide an adequate representation of an agent's uncertainty on a specific set of events of interest? More concisely and somewhat more generally: What makes a forecast $\Phi$ (ir)rational? This interrogates us about the meaning of probability and the extent to which this is captured or even reflected by its axiomatisation. Whilst the question of adequacy cannot be fully addressed independently of the context in which the uncertainty quantification is being carried out, it turns out that the logical notion of coherence is a good candidate to serve as foundation; as we now discuss.

In Chapter V of [26], a long thread of work is summed up in the observation that for probability to be useful in physics, it must bear a physical meaning. Truesdell finds it in asymptotic phenomena, i.e. those produced by a system which is composed of a very large number of elements (or a very long time series). Asymptoticity allows one to ground accurate predictions about the system solely on the system's average behaviour. Pólya [27] puts forward a similar line of thought where probability is restricted to random mass phenomena. Owing to the influence of Kinchin, this physical understanding of probability occupies centre stage in the above-recalled Kolmogorov axiomatisation, which responds in part to Hilbert's Sixth Problem. In it, probability is clearly intended as a branch of physics:

> 6. *Mathematical Treatment of the Axioms of Physics.* The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which already today mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.

See [28] for a historical appraisal.

If events – the bearers of probability – describe properties of interest of asymptotic systems, then a variety of results, which originate in Bernoulli's Law of Large Numbers *justify* the mathematics of probability as a device to predict, under specific circumstances, the future behaviour of such systems. This underpins the kind of reasoning which is invoked in inferential statistics, when data are taken to be a random sample from a hypothetical infinite population. Through this idea, **asymptotic events**, as we call them, together with the meaning they imbue to probabilistic statements, permeate all of experimental science.

However, it has long been noted that many events of interest in uncertain reasoning and forecasting fall short of asymptoticity. Either because the system has relatively few components or because it gives rise to events which are essentially unique in the sense that repetitions of the same experiment are not guaranteed to be independent. Recall that as a consequence of stochastic independence, it makes no (probabilistic) difference whether one tosses a large number $N$ of coins simultaneously or a single coin $N$ times. But as soon as one moves away from clear-cut cases where no reasonable physical interaction can be expected, arguing in favour of the independence of the components of a stochastic system can be very tricky [29].

Kolmogorov is very clear about this:

> Historically, the independence of experiments [...] represents the very mathematical concept that has given the theory of probability its peculiar stamp. We thus see, in the concept of independence, at least the germ of the peculiar type of problem in probability theory. In this book [...] we are interested mainly in the logical foundation for the specialised investigations of the theory of probability. In consequence, one of the most important problems in the philosophy of the natural sciences is – in addition to the well-known one regarding the essence of the concept of probability itself – to make precise the premises which would make it possible to regard any given real events as independent. This question, however is beyond the scope of this book. ([24], p. 8–9.)

We note in passing that what Kolmogorov means by "logical foundation" is limited to exploring the mathematical consequences of his axiomatisation.

Defending assumptions of independence is typically far from obvious in a wide class of forecasting problems spanning the social sciences (e.g. in forecasting macroeconomic variables), the life sciences (e.g. forecasting the individual patient's chance of developing a particular disease), and everyday reasoning (e.g.

forecasting the outcome of one's own wedding). This is why the aforementioned are examples of events which are only ever going to happen once. So, in contrast to asymptotic events, we shall call them **singular events**.

Forecasting singular events is central to science, technology, institutional, and personal decision-making, because even in those cases in which an experiment can be "replicated" in principle, in practice, it may be hard to be confident that they will be "close enough" repetitions of the same experiment. So it has long been felt that a central problem in the theory of probability was to provide a principled way to dispense with the mathematical comfort provided by asymptoticity.

de Finetti [30] sought to reconcile the central importance of singular events with the powerful mathematics which originates with asymptotic events. This is where the logical notion of coherence enters perhaps unexpectedly, but rather spectacularly, the stage. For as a preliminary step to achieve his grand goal, de Finetti put forward in [8] the following justification for the probabilistic quantification of uncertainty, which has become known as the *Dutch Book Argument,* see [31] for an accessible and recent overview. The central assumption in de Finetti's argument is that in a suitably defined betting problem, no rational person would willingly put themselves in a position to gamble for a *sure* loss. Doing so, de Finetti points out, would amount to obvious *incoherence.* Then he sets up a framework in which he shows that a necessary and sufficient condition to avoid this kind of incoherence is to forecast according to the prescriptions of (4) and (6). Note that (5) is easily derivable from them in standard probability logic, to which we come back later.

### 1.3.1  A logical formulation of coherence

Whilst de Finetti dubs coherence a *logical* constraint on forecasting, and indeed, the only one to qualify as such, he joins the legion of those who limit themselves to naive set theory to account for the "logic of events," and moves on to his main concern: establishing the adequacy of normalisation and finite-additivity in defining coherent forecasts. Following earlier intuitions by É Borel, he takes the probability of an event $\phi$ as the price $p \in [0, 1]$ that an individual is willing to pay in exchange for the quantity $X$ which returns 1 if $\phi$ occurs, and nothing otherwise. In this way, the identification of an individual's (rational) forecasts can be analysed within the traditional framework of *fair betting,* which goes back at least to Leibniz [32], see, e.g. [33] for a historical appraisal.

In the past few decades, research in probability logic showed that logical formulations of de Finetti's notion of coherence is not only methodologically sound [34] but also mathematically insightful [35,36]. One obvious aspect which emerges from putting coherence on a logical footing is the direct role of logical semantics in the *resolution of uncertainty*. This, as we will point out in Section 2.2, guarantees that his Dutch Book method can be extended far beyond the logic of Boolean events assumed by de Finetti, with important consequences on the pressing problems of KKR in AI, as we will point out.

So let $\phi_1, \ldots, \phi_n$ be elements of $\mathcal{SL}$ and suppose a **bookmaker** (B) publishes a forecast, or as it sometimes called in this context, *book* $\Phi : \phi_1 \mapsto a_1, \ldots, \phi_n \mapsto a_n$, where $a_1, \ldots, a_n$ are reals in $[0, 1]$. Then a **gambler** (G) chooses real-valued stakes $\rho_1, \ldots, \rho_n$ and for $i = 1, \ldots, n$, pays $\rho_i a_i$ to B. When a (Boolean) valuation $v$ resolves all uncertainty about $\phi_i$, B gains $\rho_i$, if $v(\phi_i) = 1$ and 0 otherwise. In this very special case, no one can do any better than forecasting according to Boolean logic. If some uncertainty remains unresolved, then the problem is for B to choose the $a_i$ knowing that $\rho_i$ may be negative. In this latter case, G will be paying $\rho_i a_i$, i.e. B will be receiving $-\rho_i a_i$ (similarly, receiving $\rho_i \phi_i$ means paying $-\rho_i a_i$). The resulting (abstract) gambling setting, which is indeed more articulated than it is necessary to recall here – see [23,36,37] for full details –, is aimed at providing an operational definition of what it means for bookmaker B to put forward a clearly irrational forecast, that is one leading B to a sure loss. de Finetti suggests calling $\Phi$ **coherent** if there is no choice of stakes $\rho_1, \ldots, \rho_n$ such that for every valuation $v$,

$$\sum_{i=1}^{n} \rho_i(a_i - v(\phi_i)) < 0, \tag{8}$$

i.e. if no sure loss can arise from it.

The left-hand side of (8) captures the bookmaker's payoff relative to $\Phi$, which of course depends on how the uncertainty about the events in the book are resolved by the valuation $v$.

de Finetti's Dutch Book theorem for $\mathcal{L}$-probability functions then reads as follows, see [23].

**Theorem 1.** *Let* $\phi_1, ..., \phi_n \in \mathcal{SL}$ *and let* $\Phi : \phi_i \mapsto a_i$, $i = 1, ..., n$ *be a book. The followings are equivalent*:
(1) $\Phi$ *is a coherent forecast*.
(2) *There is an* $\mathcal{L}$-*probability function, which agrees with* $\Phi$.

Coherence on singular events embodies probability with an apparently very different meaning, compared to asymptotic events, which as recalled earlier leads to probability being a property of a given physical system. It is, therefore, conceptually and mathematically very remarkable then that de Finetti was able, with his celebrated *Representation theorem*, to recover physical probability from coherence plus the weakening of independence known as *exchangeability*. This latter is introduced in [30] as a mild condition of symmetry which a set of observations enjoys if one admits that the order with they come by is irrelevant. It can be easily seen to be weaker than independence, which clearly admits no "learning" from the subsequent instances observed. Thus, events which a probability function regards as exchangeable provide viable grounds for inferring properties of the as-yet unseen instances of the hypothetical population which is producing the observations. A special case of the representation result pins down the conditions under which a probability function arising from an (infinite) exchangeable set of events is uniquely recovered as a mixture of independent and identically distributed Bernoulli random variables. We shall come back to the logical formulation of this result in Section 3, where it is seen to underpin one of the most exciting current developments in the field of pure inductive logic.

# 2 Putting logic upfront

Contemporary presentations take probability theory to be firmly rooted in measure theory. For example, David Williams puts it as follows:

> You cannot avoid measure theory: an *event* in probability is a measurable set, a *random variable* is a measurable function on the sample space, *the expectation of a random variable* is its integral with respect to the probability measure; and so on. [38] (original emphases).

This perspective is mainstream in contemporary mathematics, and certainly its fecundity lends very strong arguments to it. However, one contention of this note is that there is much to be gained from complementing it with a logical perspective on probability, and more generally on the foundations and applications of forecasting.

At least since the seminal work by Gaifman [39], who elaborates on a conjecture of Horn and Tarski [40], logicians have been interested in relating measurable sets with Boolean algebras. The problem originates with von Neumann's early work on $\sigma$-algebras dating back to the late 1930s, see [41] for a recent reconstruction of this research thread. Since Boolean algebra is (classical) logic by another name, the question arises as to which contribution does a logical foundation to probability may provide. The point is made as follows in [42]:

> Since events are always described in some language they can be identified with the sentences that describe them and the probability function can be regarded as an assignment of values to sentences. The extensive accumulated knowledge concerning formal languages makes such a project feasible.

The remainder of this note is devoted to illustrating that such a project is not only feasible, but today appears all the more promising.

## 2.1 The set of elementary events

Suppose you recently become friends with an athlete and want to form your belief that she wins her next match. Without any understanding of the sport you partition, the event space into $W, L$ and in the absence of all knowledge you assign a win the same probability as a loss. However, you might have also partitioned the event space into $W, L, D$ also allowing for the possibilities of draws and in the absence of all knowledge you assign all three outcomes the same probability, cf. [43]. Clearly, these two probability assignments are inconsistent. Some have argued that these inconsistencies spell doom to the project of applying logic to uncertain inference (references omitted to protect the guilty).

It seems more useful to us to acknowledge that some dependence of probabilities on the choice of the partition/domain is unavoidable [44]. The choice of the partition is not arbitrary but instead reflects the way you perceive the problem, i.e. what are the elementary events? Now, that providing different answers to this question leads to different probabilities should not come as a surprise to anyone. Instead, it drives home the point that how we see the world (how we formalise the problem at hand), at least partly, determines our probabilities, cf. [45].

While some dependence on the underlying domain is unavoidable, some dependencies are worse than others. Intuitively, automorphisms of the underlying fixed set of elementary events map elementary events to their doppelgangers. In the *absence of all other knowledge*, we have no logical reason to discriminate between an elementary event and its doppelganger. Hence, both ought to be assigned the same probability. Note that this argument crucially hinges on: (i) a fixed choice of an underlying domain reflecting the way you perceive the world and (ii) that there is no other knowledge. If just one of these preconditions is not met, then the argument becomes open to counter-examples. As Paris put it (Maxent is discussed in Section 3.3):

> Most of us would surely prefer modes of reasoning which we could follow blindly without being required to make much effort, ideally no effort at all. Unfortunately, Maxent is not such a paradigm; it requires us to understand the assumptions on which it is predicated and be constantly mindful of abusing them. [46, p. 6193]

Strictly speaking, it is always next to impossible to determine an exhaustive set of elementary events; e.g. the next game might be cancelled due to bad weather or a collapse of the league, annulled due to a bribed referee or the athlete might die prior to the game. To cover all these cases, a catch-all hypothesis can be added to the set of elementary events formalised as the complementary event of the union of all other elementary events, cf. [47].

## 2.2 Beyond Boolean uncertainty resolution

Since Shannon's foundation of information theory [48,49], it is a commonplace that *information resolves uncertainty*. Yet in practice it may not be clear how this happens, and which principles uncertainty resolution should follow. In informal scientific reasoning, it is often taken for granted that obtaining information means crossing out those possibilities which are inconsistent with the information obtained and represented as, say, sentences of the classical propositional calculus.

Building on this, de Finetti's analysis of coherence helped pinning down the importance of uncertainty resolution in the theory of probability, for any forecast needs to be evaluated against the realised values, as made explicit by (8). In it, de Finetti (tacitly!) assumes that the uncertainty resolution mechanism is provided by (the algebraic equivalent of) classical logic. This is convenient from the point of view of the calculus, but there are many cases in which this is not necessarily an adequate modelling choice. In those contexts, which are prominent in economics and AI, Shannon's commonplace has no straightforward, or unique application. For the way in which agents receive information is far from being unique. Good news is just waiting for us round the corner, for distinct logical semantics can lead to distinct uncertainty resolution methods.

### 2.2.1 Graded uncertainty resolution

Theorem 1 which, as anticipated, de Finetti had no inclination to cast logically, turns out to be remarkably robust to the variation of the underlying logical semantics, and hence to the associated uncertainty resolution method. A telling illustration of this point was put forward by Paris, who coined the expression *Dutch Book method* in [5] to emphasise the intrinsic generality of de Finetti's coherence-based argument. Paris pins the following "parametric" result:

**Theorem 2.** *A forecast Φ is coherent if and only if Φ is a (finite) convex combination of truth values* [5].

That of Boolean semantics (3) is then just a special case in which a logical semantics is shown to yield, via convexity, rational forecasting. Paris illustrates that by instantiating distinct notions of truth values, i.e. distinct formalisations of uncertainty resolution, a number of variations on de Finetti's notion of coherence follow from Theorem 2. Among them is the *partial* uncertainty resolution which underpins Dempster-Shafer belief functions, to which we will come back in Section 4.2.

Theorem 2 motivated further explorations of logic-based probability. Of particular note are those which took place in the field of many-valued logics, especially fuelled by Mundici's contributions [50,51]. From the point of view of our KEY PROBLEM, casting the logical foundations of coherent forecasting on many-valued logics allows us to bring to bear on uncertain reasoning a vast body of results and techniques from the theory MV-algebras. Those are to Łukasiewicz real-valued logic what Boolean algebras are to classical propositional logic, see [52,53] for classic presentations and [54] for more recent developments.

In addition to satisfying (3), Łukasiewicz real-valued logic admits further conjunction- and disjunction-like connectives, $*$ and $\oplus$, respectively, whose truth functions are defined as follows:

$$mv(\theta * \phi) = \max\{0, x + y - 1\}, \tag{9}$$

$$mv(\theta \oplus \phi) = \min\{1, x + y\}, \tag{10}$$

where $x, y \in [0, 1]$ and $+$ and $-$ are the ordinary operations on $[0, 1]$. Going back to the Fréchet bounds (2) for the probabilities of conjunctions and disjunctions, it is apparent that they arise from the combination of the corresponding Boolean and Łukasiewicz truth functions. This suggests that an interesting question in applied logic is to pin down the properties of those probability functions which are indeed compositional, see [55] for recent work on this.

Using the (algebraic equivalent) of the uncertainty resolution granted by the semantics of Łukasiewicz logic leads to the following rendering of de Finetti's argument. Let as usual $\phi_i \in \mathcal{SL}$ be events of interests, for which a bookmaker $B$ writes a forecast $\Phi : \phi_i \mapsto a_i \in [0, 1]$. Gambler $G$ chooses stakes and pays bookmaker $B$, for each $\phi_i$, $\sigma_i \cdot a$, while $G$ receives from $B$, $\sigma_i \cdot mv(\phi_i)$, that is an amount proportional to the degree of truth of $\phi_i$. A forecast $\Phi : \phi_i \mapsto a_i$ is **MV-coherent** if it is not the case that for every Łukasiewicz $[0, 1]$-truth valuation $mv$,

$$\sum_{i=1}^{n} \sigma_i(a_i - mv(\phi_i)) < 0. \tag{11}$$

It is shown by [51] that a forecast on an MV-algebra $\mathbf{M}\Phi : \phi_j \mapsto a_j$ is MV-coherent if and only if $\beta$ extends to a *state*, i.e. a finitely additive and normalised measure on $\mathbf{M}$, see [54] for details on the theory of states.

In addition to advancing the logico-algebraic foundations of probability, which includes work on the notion of *strict coherence* [37,56], the many-valued extension of the Dutch Book Method plays an important role in bringing graded methods and techniques from mathematical fuzzy logic [57] to uncertainty in AI [4].

Related to fuzzy logic, but independent of coherence, is the "two-layered" approach which takes probability as a graded *modal logic*. The key idea put forward in [58,59], is to use Łukasiewicz real-valued semantics to define $P(\phi)$ as the degree to which the modal sentence "probably $\phi$" is true. This approach has recently been shown in [60] to be essentially equivalent to the probabilistic logic put forward in [61], and to which we will come back in Section 4.

### 2.2.2 Partial uncertainty resolution

A second noteworthy variation on the theme of Boolean uncertainty resolution adds interesting features to the logical approach to our KEY PROBLEM, namely, relaxing the hypothesis to the effect that all uncertainty is fully resolvable in de Finetti's setup. Early contributions to this problem can be found in microeconomics, and in particular in the work of Jaffray [62]. In logical terms, the idea here is that uncertainty is resolved through the (classical) logical closure of the information possessed by an agent, rather than through a (Boolean) valuation which decides *all* events of interest at once. Hence, if a given event $\phi \in \mathcal{SL}$ is known to have occurred, the only uncertainty that can be (logically resolved) pertains every (satisfiable) $\theta \in \mathcal{SL}$ such that $\phi \vDash \theta$. Conversely, there may be events of interest for which no uncertainty can be resolved at all.

This leads to the following generalisation of de Finetti's method. Bookmaker $B$ publishes $\Phi : \phi_i \mapsto a_i$ and then $G$ places stakes $\sigma_1, \ldots, \sigma_k$ on $\phi_1, \ldots, \phi_k$, i.e. at the prices written in $\Phi$. Finally, $G$ pays, for each $\phi_i$ the amount $\sigma_i \cdot a_i$ to $B$. Hence, $B$ will gain the amount $\sigma_i \cdot C_\phi(\phi_i)$, where $C_\phi(\phi_i) = 1$ if $\phi \vDash \phi_i$, and $C_\phi(\phi_i) = 0$ otherwise. The following variation on de Finetti's definition of coherence, as captured by (8), then arises. Say that a forecast $\Phi$ is **coherent under partly resolved uncertainty** if it is not the case that, for a given satisfiable $\phi$,

$$\sum_{i=1}^{n} \sigma_i(a_i - C_\phi(\phi_i)) < 0. \tag{12}$$

The early result obtained by Jaffray, and encompassed by Theorem 2, is to the effect that a book $\Phi$ under partly resolving uncertainty is coherent if and only if it can be extended to a Dempster-Shafer *belief function* [63] on the Boolean algebra generated by $\Phi$. As we briefly recall in the next subsection, belief functions constitute a prominent non-additive approach to the quantification of uncertainty in statistics and AI.

The graded and partly resolving uncertainty resolution generalisations of de Finetti's method have been shown to be compatible by [64], where Paris's generalisation of Jaffrays' framework is formalised in logico-algebraic terms and is shown to yield a Dutch Book argument for coherent forecasts under both graded and partly resolvable uncertainty. This pins down Dempster-Shafer belief functions on *many-valued events*, i.e. events formalised in some many-valued logic, typically the Łukasiewicz real-valued one.

### 2.2.3 Connections with non-additive measures of uncertainty

As illustrated earlier in connection to the Fréchet inequalitites (2), probability intervals arise naturally if we follow Boole in seeking for a general method to assign probabilities to arbitrary events based on known probabilities. However, if forecasting is aimed at decision-making, the lack of a unique number summarising all the relevant uncertainty can raise notable difficulties. Keynes, an expert on Boole's *Laws of Thought*, suggested that we should bite the bullet:

> The sense in which I am using the term ["uncertain" knowledge] is that in which the prospect of a European war is uncertain [...] About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed [65].

Much recent work at the intersection of decision-theory, statistics and AI have been moving along the quest for more principled solutions, as illustrated by the Jaffray framework. The common trait of this approach is to relax the additivity of the uncertainty measure. How this arises is easily seen by recalling a simple problem which has been made popular by Ellsberg [66]. Consider an urn with red, blue, and green balls. Suppose $\psi_1$ stands for "the ball is red." Suppose further that the agent knows that the proportion of the red balls in the urn is 1/3. This leads to the straightforward quantification of the agent's uncertainty, $P(\psi) = 1/3$. Note however that the information available does not resolve *enough* uncertainty to allow the agent to come up with equally straightforward quantifications for either blue ($\psi_2$) or green ($\psi_3$) balls – any value in [0, 2/3] will be consistent with the information available. In the absence of any further information, the probabilistic setting would lead

to setting both $P(\psi_2)$ and $P(\psi_3)$ to 1/3. The problem with that, has been argued, is that 1/3 fails to represent the information to the effect that the available information does not support 1/3 any more than the *very many* choices in [0, 2/3]. Dispensing with additivity, Shafer [63] belief functions have been argued to provide a satisfaction quantification of uncertainty in cases of this sort, see [67] for a comprehensive overview. Belief functions occupy a central place in the multifarious theories of imprecise probabilities [68]. Recent work attempting at a logical analysis of non-additive measures of uncertainty include [69–72].

# 3 Inductive logic

This section is devoted to logical approaches for pinning down a unique probability of singular events in case the available information does not uniquely determine a probability. Having discussed the foundations in the previous two sections, we now focus on key notions and technical results.

## 3.1 Carnap

Rudolf Carnap is widely considered to be the founder of **inductive logic**. He distinguishes two concepts of probability [73] and [74, Chapter 2] (see also [75])

> Probability$_1$ is a logical concept, a certain logical relation between two sentences (or, alternatively, between two propositions); it is the same as the concept of degree of confirmation. [...] On the other hand, probability$_2$ is an empirical concept; it is the relative frequency in the long run of one property with respect to another. [76, p. 72]

He mostly studied the former concept of probability, which is interpreted as a degree of confirmation, $c$. The degree to which the evidence $e$ confirms a hypothesis $e$ is then $c(h, e)$. In our notation, $c(h, e)$ is written as $P(\theta|\phi)$. The ultimate goal of his approach is to find a probability function capturing the logical degree of confirmation.

To this end, Carnap considered a fixed first-order predicate language $L$ containing finitely many unary relation symbols (predicates), $R_1,...,R_q$, and countably many constants, $a_1, a_2,...$, as names for all possible elements of an underlying domain with a countable (possibly finite) size. The language does not contain relation symbols of higher arity, nor a symbol for equality and no function symbols. The goal is thus to find a probability function capturing the degree of confirmation a sentence $\phi$ (the evidence) bestows on another sentence $\theta$ (the hypothesis). This, again, can be seen to be a special case of our initial Key Problem.

In it, probabilities are constrained by a number of axioms such as:

**Constant exchangeability (Ex)**
The probability of a sentence does not change, if we replace the distinct constants appearing in it by any other set of distinct constants of the language.
**Johnson's sufficientness postulate (JSP)**
The probability of an unseen instance having certain properties given a sample of observations only depends on the size of the sample and the number of instances in this sample having exactly these properties.

Ex means that the way we name elements does not matter, and it is clearly a logical way of casting de Finetti's notion of exchangeability recalled above. JSP says that $P\big(a(a_{n+1})|\bigwedge_{i=1}^n a_{h_i}(a_i)\big)$ only depends on $n$ (the sample size) and the number of $h_i$ such that $\alpha = \alpha_{h_i}$, where all $\alpha$ are atoms of the language: conjunctions of the form $\alpha(a_k) = \pm R_1 a_k \wedge \pm R_2 a_k \wedge ... \pm R_q a_k$ determining all properties of a constant $a_k$.

From these two and further axioms, Carnap showed that for all languages $L$ containing at least two unary relation symbols ($q \geq 2$) all probability functions defined over $L$ satisfying this set of axioms are characterised by a one-parameter family of probability functions, $P_\lambda^L$, with $0 \leq \lambda \leq \infty$

$$P_\lambda^L\left(a_j(a_{n+1})|\bigwedge_{i=1}^n a_{h_i}(a_i)\right) = \frac{|\{i : h_i = j\}|+\lambda \cdot 2^{-q}}{n + \lambda}. \tag{13}$$

The parameter $\lambda$ can be interpreted as a measure of how quickly evidence causes probabilities to move away from the uniform probability function. If $\lambda = \infty$, then $P_\infty^L(a_j(a_{n+1})|\bigwedge_{i=1}^n a_{h_i}(a_i)) = \frac{|\{i : h_i = j\}| + \infty \cdot 2^{-q}}{n + \infty} = 2^{-q}$ for all atoms $a_j$: $P_\infty^L$ is the uniform distribution over $L$. That is, for all evidence sentences, one obtains the uniform distribution. $\lambda = 0$ means that all constants look the same with probability one: $P_0^L(a_j(a_{n+1})|a_{h_1}(a_1)) = 1$, if and only if $j = h_1$. That is, after observing some contingent fact about a constant $a$, $\phi(a)$, $P_0^L$ assigns probability one that the same fact holds for all other constants, $P_0^L(\forall x \phi(x)|\phi(a)) = 1$.

As we have seen in Section 2.1, dependence of probabilities on the underlying domain can cause serious issues. Consider a sentence $\phi$ of the language $L$. Note that $\phi$ is also a sentence of all other languages $L'$ that contain the relation symbols $R_1, R_2, ..., R_q$ and further unary relation symbols. Fortunately, Carnap's $P_\lambda^L$ do not depend on the underlying unary language $L$: for all fixed $\lambda$, all languages $L$, all sentences $\phi$ of $L$ and all such languages $L'$ it holds that $P_\lambda^L(\phi) = P_\lambda^{L'}(\phi)$ [77, Corollary 16.2].

Carnap and others tried to determine a unique value of $\lambda$ that would govern all scientific inference. No such value was found. Today, we might interpret $\lambda$ as an evidential entrenchment (or stubbornness) parameter, which represents the weight of evidence required to shift away from the prior probability distribution. The particular value of $\lambda$ or its range would then be determined by a particular application and could thus vary from case to case. Not being able to determine a unique value for $\lambda$ is from a modern perspective an advantage rather than a fatal flaw. Historians and philosophers are still grappling with Carnap's legacy [45,78–82], see also the extensive [83].

## 3.2 Pure inductive logic

Pure inductive logic is the study of rational probability from a mathematical perspective continuing Carnap's work. The underlying first order predicate language is now allowed to contain relation symbols of arbitrary arity [77]. *Pure* here refers to the convention that there is no interpretation of the non-logical symbols of the underlying language. That is, probabilities are assigned in the absence of all knowledge. The aim "is to investigate this process of assigning logical, as opposed to statistical, probabilities" [77, p. 4]. Extensions of Pure Inductive Logic to include a symbol for equality [84] or function symbols [85] are still in their infancy.

The main methodology is to postulates axioms, which putatively capture (aspects of) rationality, and then investigate the logical relationships between different sets of axioms. These axioms are normally motivated by intuitions we have about symmetry, irrelevance, relevance, and independence.

If the language contains at least one relation symbol that is not unary, $R$ say, then there are infinitely many literals required to pin down all properties of a constant $a$: for a binary $R$, one would need to write $\bigwedge_{i=1}^\infty \pm Raa_i$. This expression however is not a sentence of a first-order language, which only allows for sentences of finite length. Instead, one considers state descriptions $\alpha$ on $n$, which determine all properties of the first $n$ constants: $\alpha = \bigwedge_{R \in L} \bigwedge_{i_1=1}^n \wedge ... \wedge \bigwedge_{i_{arity(R)}=1}^n \pm Ra_{i_1} ... a_{i_{arity(R)}}$. Gaifman's seminal theorem shows that a probability function on such a language is uniquely defined by the probabilities it assigns to the state descriptions [86]. In other words, specifying probabilities for all state descriptions defines a unique probability function on the entire language.

Two constants $a, b$ are said to be equivalent over a state description $\alpha$, if and only if the sentence $\alpha \wedge a \equiv b$ is consistent (note that $\alpha \wedge a \equiv b$ is not a sentence of the language). This induces an equivalence relation $S(\alpha)$ over the first $n$ constants. The spectrum of $\alpha$ is then defined as the multiset of sizes of the equivalence relation $S(\alpha)$. In other words, constants $a$ and $b$ are equivalent, if and only if replacing all occurrences of $a$ by $b$ in $\alpha$ yields a consistent sentence and replacing all occurrences of $b$ by $a$ in $\alpha$ yields a consistent sentence. We can now order the equivalence classes by size, starting with the largest class first. The spectrum of $\alpha$ is then defined as the vector of these sizes.

This notion of a spectrum is a key in important axioms of pure inductive logic:

**Spectrum exchangeability (Sx)**
If two state descriptions on $n$ have the same spectrum, then they have the same probability.

**Language invariance with P (Li with P)**
A probability function on language $L$ satisfies (Li with P), if there is a family of probability functions $P^{\mathcal{L}}$, one for each language $\mathcal{L}$, such that $P^L = P$ and whenever $L$ is a sublanguage of $L'$, then $P^L$ and $P^{L'}$ agree on all the sentences of all $L$.
**Principle of induction (PI)**
Let two state descriptions $\beta, \gamma$ on $n + 1$ both entail a state description $\alpha$ on $n$. $P(\beta|\alpha) \geq P(\gamma|\alpha)$, if there are at least as many constants equivalent to $a_{n+1}$ over $\beta$ as there are over $\gamma$.

We normally consider a special case of (Li with P); we are mainly interested in probability functions satisfying (Sx). Thus, (Li with P) becomes (Li with Sx).

It is often hard to show directly that such (sets of) axioms have (un-)desirable consequences. A convenient tool are representation theorems in the tradition of de Finetti [87] of the form: All probability functions $P$ satisfying a set of axioms have the following form $P = \int_{\vec{p} \in \mathbb{B}} P^{\vec{p}} d\mu(\vec{p})$, where the $P^{\vec{p}}$ are probability functions generated via sampling coloured balls from an urn, and the integral is over Borel subsets of the space of possible urns. Technically, one can often show that all $P^{\vec{p}}$ have some property and that taking convex mixtures preserves this property.

For example, a probability functions satisfying (Sx) is *t-heterogeneous*, if and only if $\lim_{n \to \infty} \sum_{\alpha \in SD_n^t} P(\alpha) = 1$, where $SD_n^t$ is the set of state descriptions on $n$ such that $S(\alpha)$ has exactly $t$ classes. A probability function satisfying (Sx) is said to be *homogeneous*, if and only if $\lim_{n \to \infty} \sum_{\alpha \in SD_n^t} P(\alpha) = 0$ for all $t \in \mathbb{N}$.

de Finetti representation theorems have been provided for $t$-heterogeneous probability functions [77, Theorem 31.11], homogeneous probability functions [77, Theorem 31.10], probability functions satisfying (Sx) [77, Theorem 31.12], (Li with Sx) [77, Theorem 32.1], as well as probability functions on languages containing only unary relation symbols satisfying (Ex) [77, Theorem 9.1]. Employing these results, it was proved that
(1) All homogeneous probability functions satisfy (Li with Sx) [77, Proposition 30.1].
(2) (Li with Sx) entails (PI) [77, Theorem 36.1].
(3) Hence, all homogeneous probability functions satisfy (PI).

It is not known whether all heterogeneous probability functions satisfy (PI), this conjecture is still open [77, p. 269]. If this conjecture is true, then (Sx) entails (PI).

The status of Carnap's continuum is not so clear for languages containing relation symbols of greater arity. If $L$ contains at least one non-unary relation symbol, then there are only two probability functions consistent with straight-forward generalisation of (JSP) and (Ex): the uniform probability function and the natural analogue of $c_0$ on such languages [88] and [89, Theorem 20]. A slightly different generalisation of (JSP) is the binary sufficient-ness postulate (BSP). (BSP), (Ex), and the requirement to assign strictly positive probability to all quantifier-free sentences jointly produce an analogue to Carnap's continuum, which is parameterised by two(!) parameters [90].

After adding a symbol for equality to the underlying language, one can canonically generalise JSP and obtain (JSP_=). A probability function on such a language satisfies (JSP_=) and (Ex), if and only if it is part of a one-parameter family characterised by $0 \leq \lambda \leq \infty$. These probability functions are obtained from Carnap's $P_\lambda^L$ by, in some sense, sending the number of relation symbols in $L$ to infinity [77, Corollary 38.4]. Since these new functions satisfy spectrum exchangeability on the predicate language enriched with a symbol for equality [77, Lemma 38.2], they define a family of probability functions on the languages without equality that satisfy (Li with Sx) [77, Theorem 37.1]. Unfortunately, our understanding of how these probability functions behave on predicate languages with non-unary relation symbols is still a mystery [77, p. 289]. A more general class of probability functions containing these probability functions has been studied in [91,92]; connections to Ewens sampling are traced in [93].

Important open problems in pure inductive logic were recently surveyed in [94].

## 3.3 Inference processes

Consider an agent having acquired some information and subsequently rejected some probability functions as incompatible with the available information. For example, the agent knows the current temperature is 20 ˚C

and hence does not want to assign a probability greater than 1% to the event that it will be snowing tomorrow. Which remaining probability function(s) should the agent use for rational uncertain inference? This way of formulating our Key Problem permits us to study inductive logic by investigating the choice of a probability function from a set of exogenously given probability functions.

Formally, denote this set of remaining probability functions by $\mathbb{E}$. Given $\mathbb{E}$ and no other information, which probability function should an agent adopt? The key idea is to implement this choice via a function $f$. This function $f$ assigns every given set of probability functions a single probability function. It is natural to require that the function $f$ picks out a probability function in $\mathbb{E}$, $f(\mathbb{E}) \in \mathbb{E}$.

The, maybe, most basic inference process on a finite set of elementary events $\Omega = \{\omega_1, ..., \omega_n\}$ is given by the function $f$ which maps $\mathbb{E}$ to its centre of mass (CM). This choice is, in some sense, an average choice. While this choice seems reasonable at first glance, the probabilities picked out by $f_{CM}$ depend on the underlying domain in an unfortunate sense. Adding a further variable $Z$, one obtains an enlarged domain $\Omega^+ := \{\omega_1 \wedge z, \omega_1 \wedge \neg z, ..., \omega_n \wedge z, \omega_n \wedge \neg z\}$. Supposing we have no information about $Z$ denote by $\mathbb{E}^1$ the set of probability functions on $\Omega^+$ consistent with $\mathbb{E}$. The highly desirable axiom of Language invariance (LI) now demands that $f_{CM}(\mathbb{E}^1){\restriction}_\Omega = f_{CM}(\mathbb{E})$. CM does not satisfy LI [95, p. 73]. For example, the CM probabilities for rain tomorrow may change, if a variable for the bread prices in Venice in the year 3456 is included in the underlying domain and we are, obviously, ignorant about these prices.

The obvious way to enforce language invariance in the spirit of CM is to add more and more such variables $Z_1, Z_2, ...$ and define $f_{CM}^\infty(\mathbb{E}) := \lim_{n \to \infty} f_{CM}^n(\mathbb{E}^n){\restriction}_\Omega$, where $f_{CM}^n(\mathbb{E}^n)$ is the centre of mass of $\mathbb{E}^n$. This construction gives rise to the inference process $CM_\infty$, which satisfies language invariance by construction [96, Theorem 5]:

$$f_{CM}^\infty(\mathbb{E}) = \arg\sup_{P \in \mathbb{E}} \sum_{\substack{\omega \in \Omega \\ \exists Q \in \mathbb{E}:\, Q(\omega) > 0}} \log(P(\omega)).$$

Assuming that $\mathbb{E}$ is convex, $f_{CM}^\infty(\mathbb{E})$ is the point in $\mathbb{E}$ that is, in some sense, closest to the uniform probability function, $P_=$. In case $P_= \notin \mathbb{E}$, $f_{CM}^\infty(\mathbb{E})$ is in the boundary of $\mathbb{E}$. $CM_\infty$ is thus as far away from the original motivation of CM to pick a probability that are as average or typical as possible [95, pp. 73–76].

The most studied inference process is MaxEnt arising from maximising Shannon entropy of a probability function [48], $H(P)$, on convex sets $\mathbb{E}$:

$$f_{ME}(\mathbb{E}) = \arg\sup_{P \in \mathbb{E}} H(P) = \arg\sup_{P \in \mathbb{E}} - \sum_{\omega \in \Omega} P(\omega) \cdot \log(P(\omega)),$$

with the usual convention that $0 \cdot \log(0) := 0$.

Vencovská and Paris provided axiomatic characterisations of ME inference in terms of the so-called common sense principles of reasoning [95,97–101]. For example, probabilities of interest ought not to depend on irrelevant information, should be invariant under renaming and should not change unless evidence to the contrary is received. Importantly, ME is language invariant [95, Section 6]. See [102] for more discussion on CM and MaxEnt. Furthermore, ME is invariant under automorphisms [46] in a broader sense than discussed in Section 2.1.

Another type of axiomatisation harkens back to Boltzmann. In 1871, he showed that the micro-states of gas molecules are most likely very close to a maximum entropy state. It is demonstrated in [103], that on very large domains almost all probability functions consistent with certain types of constraints are very close to the maximum entropy probability function.

The explication of the notion of information by Shannon entropy has given rise to a great number of axiomatisations. A very good, although by now dated, overview is [104] delineating different types of axiomatisations. Further classical key contributions were made in [105–108]. It is shown in [109] that the sort of principles which characterise maximum entropy reasoning guarantee a very high degree of conformity in the choices made by distinct agents facing essentially the same choice problem.

Further strands of research (i) characterise the maximum entropy function as a unique solution to minimax optimisation problems. These problems are motivated by decision theoretical or information theoretical considerations [110–112]; (ii) study generalised notions of maximised entropy [110,113–116]; (iii) consider multi-agent settings [117] and investigate probabilities of conditionals [118–120].

## 3.4 Inference processes on infinite domains

Inference processes on infinite domains are rarely studied in generality [44]. Almost all these works focus on a maximisation of Shannon Entropy, one exception is [121]. There are, however, three obstacles for defining entropy maximisation on an infinite domain Ω by simply letting

$$H(P) := - \int_{\omega \in \Omega} P(\omega) \cdot \log(P(\omega)) \mathrm{d}\omega.$$

(1) It is not clear which measure $\mathrm{d}\omega$ to adopt. (2) There will be many cases with infinitely many probability functions with infinite entropy and hence no maximal entropy function. (3) On countable domains Ω, there exists no uniform probability function and hence no intuitive probability function to adopt in the absence of evidence.

Two explications of the maximum entropy principle have been put forward, if the domain is given by an underlying first-order language as it is in pure inductive logic. The first explication employs limits of maximum entropy functions on finite sublanguages [122,123]. Calculating these limits is an obvious way to construct the probabilities for inference. There are unfortunately cases in which these limits are ill-defined [124], and in other cases, the constructed probabilities do not satisfy the given premises [125]. The second approach employs a pre-order on the set of probability functions consistent with the evidence. This pre-order explicates a *greater-entropy-than* relation. The probabilities for inference are provided by the maximal elements of the pre-order. This second approach is less constructive but more widely applicable [126–130]. It has been conjectured that these two approaches agree where the former explication is well-defined; the conjecture has so-far been verified only in a number of cases [131].

# 4 Applications to AI

Researchers in AI have long understood that they require tools other than classical logic to represent and reason with information [132]. Approaches and applications for uncertain inference abound. A relatively dated but still very useful introductory review to the use of probability logic in AI is available in [133]. Darwiche [134] pins down the three roles which logic is expected to have in the coming of age of the new AI. The aforementioned handbook [4] provides a very useful overview of the current use of logic in the subfield of AI which goes under the heading of KRR.

The growth of AI and AI applications has been spectacular over the last few decades. The field of AI, if one may still speak of AI as a single field, has become so large, that we cannot event attempt to give an overview of logic for uncertainty in AI. In the following, we focus on a few select areas, which have close connections to the topics discussed earlier.

## 4.1 Bayesian networks

Bayesian networks are a graphical tool to represent (conditional) independences of probability functions. The possibility to calculate some conditional probabilities by considering only (small) parts of the network and to organise information graphically have contributed to their proliferation [135,136]. The possibility to effectively compute probabilities make Bayesian networks well-suited for producing probabilities for forecasting.

Formally, Bayesian networks consist of a directed acyclic graph defined on a finite set of variables $V = \{X_1, …, X_n\}$. A variable $Y$ is a parent of a variable $X$, if and only if a directed edge points from $Y$ to $X$. The second part of the Bayesian network is the specification of (conditional) probabilities for all variables $X \in V$ relative to the set of the parents of $X$, $\mathrm{Par}_X = \{X_1, …, X_{g_X}\}$ and all the possible values the parent variables can take:

$$P(X = x | X_1 = x_1, ..., X_{g_X} = x_{g_X}).$$

To ensure that a probability function is defined it needs to hold that for all $X \in V$ and all possible values of the parent variables $X$ that $1 = \sum_{x \in X} P(X = x | X_1 = x_1, ..., X_{g_X} = x_{g_X})$. The absence of directed edges in a Bayesian network entails probabilistic independences of variables. This property facilitates the communication of modelling assumptions, information stored in data and/or knowledge bases between authors, readers, (software) engineers, and referees.

While worst-case complexity of inference in Bayesian networks is #P-complete [137] ("the answer to a #P-complete problem is the number of solutions to some NP-complete problem" [136, p. 170]), there has been much work on approximating the correct results and finding tractable sub-classes. Returning to themes of Section 3, Bayesian networks have also found applications for MaxEnt reasoning [138–140]. More generally, learning graphical models representing knowledge or databases has been of considerable interest [141].

Bayesian networks are also widely successful in modelling **causal** reasoning. Let us consider this example: The road, $R$, could either be wet be**cause** of last night's weather, $W$, or be**cause** a sprinkler was on last night, $S$. So, $R = $ wet, if and only if $W = $ rain or $S = $ on. The DAG structure of the Bayesian network is shown in Figure 1. Initially, last night's weather and the sprinkler are independent. They become dependent once it is observed that the road is wet: Knowing that it did not rain last night entails that the sprinkler was on. However, *intervening* on the condition of the road by sending out a cleaning and drying machine, renders the road variable $R$ independent of the other two variables. In Pearl's *do calculus*, intervention on variables (here $R$), means that all edges in the DAG pointing to this variables are deleted [142]. Hence, $W$ and $S$ are independent given do($R = $ dry). Summarising, $W$ and $S$ are initially independent, dependent (correlated) after observing $R$ but independent after intervening on $R$.

Finally, not only might the relationships between variables be non-deterministic one may also be uncertain about the underlying causal structure. The underlying structure may be partially learnt from experts [143], data [144,145], which has certain theoretical limits [146], or be expressed as Bayesian probabilities attaching to causal Bayesian networks [147]. In the latter case, evidence concerning the causal structure is incorporated via Bayesian/Jeffrey updating, for more references on Jeffrey updating see [148].

Albeit less popular, further graphical models for uncertain reasoning have found applications. Credal nets also employ directed acyclic graphs but, unlike Bayesian networks, the conditional independences apply to a set of probability functions [149,150]. Markov nets instead employ undirected graphs and a single probability function. See [151] for an overview of further graphical models and an overview of applications.

## 4.2 Logics for reasoning about probabilities

As we laid out earlier, probability functions assigning sentences of a logical language some probability in the unit interval $[0, 1]$ are natural extensions classical valuations, which assign sentences values in a set of size two {false, true} or equivalently in the set $\{0, 1\}$. Such an approach does not permit the comparison or a Boolean combination of probability statements to be formalised within the logical language. Statements of the form $P(\phi) < P(\theta)$ and $P(\phi) + P(\theta) = 0.3$ have to be evaluated at a meta-level or be encoded by using sets of probabilities, see Section 2.2. Another approach to formalise the remaining uncertainty is to include probabilistic operators in the language.
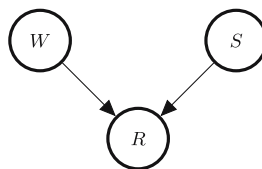


**Figure 1:** Directed acyclic graph of the Bayesian network for the sprinkler example.

There is no unique canonical approach that incorporates such statements as part of the logic, because there are number of choices to be made, which do not have an obviously correct option. One may chose

(1) a logical language (propositional, first-order logic, second-order logic, fragments thereof, etc.),

(2) a logic (classical, modal [152,153], many-valued, temporal, etc.),

(3) a representation of uncertainty (probabilistic, imprecise probabilities, ranking functions, partially ordered, containing infinitesimals, etc.),

(4) a set of operators for combining uncertainties (comparisons of uncertainties, uncertainties equalling some external value(s), combination of uncertainties (e.g. linear combinations $P(\phi) + 2 \cdot P(\theta) = 0.3$)),

(5) whether knowledge and uncertainty's fundamental representation is conditional ([35,119,120,154]).

As mentioned earlier, now classical works in this area include [61,155], see also [156] for a comprehensive overview of this approach. These works study classical properties of such logics such as decidability, computational complexities, soundness, and completeness. A group in Belgrade has been active in this area, e.g. [19,157].

## 4.3 The future of logic and probability for AI

Every formal model of learning, including machine learning techniques and neural network models, must make some assumptions concerning underlying regularities of the domain of interest. For, if the world is without regularities, then all previous observations have no bearing on (predictions of) the future. The necessity of bias supplied by humans for successful prediction is accepted in modern-day AI and was anticipated in the philosophy of science by decades:

> In constructing an induction machine we, the architects of the machine, must decide *a priori* what constitutes its "world"; what things are to be taken as similar or equal; and what *kind* of "laws" we wish the machine to be able to "discover" in its "world." In other words we must build into the machine a framework determining what is relevant or interesting in its world: the machine will have its "inborn" selection principles. The problems of similarity will have been solved for it by its makers who thus have interpreted the "world" for the machine. [158, p. 48]

Logics, such as inductive logics, can help by (i) providing biases for learning, (ii) providing means for analysing biases in successful learning systems, and (iii) elucidate black box prediction systems (such as deep neural networks) by making predictions more transparent and explainable. Graphical tools, such as Bayesian networks, can also help expert users (e.g. professionals) explain decisions and decision procedures of AI systems to lay people [159].

Our ability to create, store and manipulate ever larger data sets has led some to believe that Big Data can be the solution to many of our inference problems by *letting the data speak for themselves*. The success of data-driven methods has indeed been spectacular. If this hypothesis were widely true, then there would be no more need for (logically) approaches to uncertainty; data crunching would suffice.

However, data alone is never enough – not even for asymptotic events:

> The data *cannot* [emphasis original] speak for themselves; and they never have, in any real problem of inference.
> For example, Fisher advocated the method of maximum likelihood for estimating a parameter; in a sense, this is the value that is indicated most strongly by the data alone. But that takes note of only one of the factors that probability theory (and common sense) requires. For, if we do not supplement the maximum likelihood method with some prior information about which hypotheses we shall consider reasonable, then it will always lead us inexorably to favor the "sure thing" hypothesis $H_S$, according to which every tiny detail of the data was inevitable; nothing else could possibly have happened. For the data always have a much higher probability (namely $P(D|H_S) = 1$), on $H_S$ than on any other hypothesis; $H_S$ is always the maximum likelihood solution over the class of all hypotheses. Only our extremely low prior probability for $H_S$ can justify our rejecting it. [1, p. 195]

For more recent arguments about the untenability of purely data-driven forecasting, see [160,161].

Despite initial appearances to the contrary, logical approaches to uncertain reasoning far from being out-of-date are rather gaining new actuality in the face of fast developing machine learning and AI, more widely as a foundational approach underpinning and support cutting-edge research.

# 5 Conclusion

In this note, we highlighted connections between logic and probability as underpinnings for reasoning under uncertainty. Two main strands of active and fruitful research have been illustrated: pure research and advances in practical applications. It seems appropriate to conclude with a forecast on how they will develop in the medium term.

On the pure side, we can only expect probability theory to advance the current borders of mathematical logic. One further area which has been investigated with some success in this spirit is to do with the existence of 0-1-laws in several logics. Informally, a logic has a 0-1-law if for each sentence $\theta$ of the logic, the asymptotic probability of $\theta$ is either 0 or one. 0-1-laws have now been established for a number of logics [127,162–164].

The rapid pace with which AI in general and machine learning have been advancing in the past decade shows no signs of slowing down. But we can now take for granted that big data will not be the silver bullet giving us the solution to all our inference problems – we still require explanations for opaque black box predictions, which can be supplied by combining logic and probability [159,165]. We also need to provide an inductive bias to make learning possible (Section 4.3).

We think that much has been gained from intertwining logic and probability, and forecast that this will remain true in the future.

**Conflict of interest**: The authors state no conflict of interest.

**Data availability statement**: Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

# References

[1]     E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.
[2]     J. van Benthem, *Logic and reasoning: Do the facts matter?*, Studia Logica **88** (2008), no. 1, 67–84.
[3]     B. de Finetti, *The logic of probability (1935)*, Philos. Stud. **77** (1935), 181–1990.
[4]     P. Marquis, O. Papini, and H. Prade, *A Guided Tour of Artificial Intelligence Research*. Vol. 1: *Knowledge Representation, Reasoning and Learning*, Springer, Cham, 2020.
[5]     J. B. Paris, *A note on the Dutch book method*, in: G. De Cooman, T. L. Fine, T. Seidenfeld (Eds.), ISIPTA '01: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications (Ithaca, NY, USA), Shaker Publishing B.V., 2001, pp. 301–306.
[6]     A. Hald, *History of Probability and Statistics and Their Applications before 1750*, Wiley, Hoboken, 1990.
[7]     B. de Finetti, *Sul significato soggettivo della probabilità*, Fund. Math. **17** (1931), 289–329.
[8]     F. P. Ramsey, *Truth and probability (1926)*, in: R. B. Braithwaite (Ed.), The Foundations of Mathematics and other Logical Essays, Routledge & Kegan Paul, London, 1931, pp. 156–198.
[9]     B. de Finetti, *Teoria delle probabilità*, Einaudi, 1970.
[10]    L. J. Savage, *The Foundations of Statistics*, *Dover Publications*, 2nd ed., New York, 1972.
[11]    R. C. Jeffrey, *Bayesianism with a human face*, Testing Scientific Theories Minnesota Stud. Philosophy Sci. **10** (1983), 133–156.
[12]    I. Hacking, *The Emergence of Probability*, Cambridge University Press, Cambridge, 1975.
[13]    T. Hailperin, *Probability logic*, Notre Dame J. Form. Log. **25** (1984), no. 3, 322–332.
[14]    T. Hailperin, *Sentential Probability Logic. Origins, Development, Current Status, and Technical Applications*, Academic Press, Bethlehem, 1996.

[15] A. De Morgan, *Formal Logic*, Taylor, London, 1847.

[16] G. Boole, *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*, Walton and Maberly, London, 1854.

[17] T. Hailperin, *Best possible inequalities for the probability of a logical function of events*, Amer. Math. Monthly **72** (1965), no. 4, 343–359.

[18] W. Harper and C. A. Hooker, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, University of Western Ontario, 1976.

[19] Š. Dautović, D. Doder, and Z. Ognjanović, *An epistemic probabilistic logic with conditional probabilities*, in: W. Faber, G. Friedrich, M. Gebser, M. Morak (Eds.), Logics in Artificial Intelligence, JELIA 2021, Lecture Notes in Computer Science, Vol. 12678, Springer, Cham, 2021, pp. 279–293.

[20] T. Flaminio, L. Godo, and H. Hosni, *Boolean algebras of conditionals, probability and logic*, Artificial Intelligence **286** (2020), 103347.

[21] V. S. Vladimirov, *Boolean Algebras in Analysis Mathematics and Its Applications*, Kluwer Academic Publishers, Dordrecht, 2002.

[22] E. Borel, *Probability and Certainty*, Walker and Company, New York, 1950.

[23] J. B. Paris, *The Uncertain Reasoner's Companion: A Mathematical Perspective*, Cambridge University Press, Cambridge, 1994.

[24] A. Kolmogorov, *Foundations Theory of Probability*, Chelsea Publishing Company, New York, 1950.

[25] D. Dubois, H. Prade, and R. Sabbadin, *Decision-theoretic foundations of qualitative possibility theory*, European J. Oper. Res. **128** (2001), no. 3, 459–478.

[26] C. A. Truesdell, *Six Lectures on Modern Natural Philosophy*, Springer, Berlin, 1966.

[27] G. Pólya, *Mathematics of Plausible Reasoning*, Vol. 1-2, Princeton University Press, Princeton, 1954.

[28] L. Corry, *David Hilbert and the Axiomatization of Physics (1898–1918) – From Grundlagen der Geometrie to Grundlagen der Physik*, Springer, Dordrecht, 2004.

[29] M. Kac, *Statistical independence in probability, analysis and number theory*, The Mathematical Association of America, John Wiley and Sons, Inc., 1959.

[30] B. de Finetti, *La prévision: ses lois logiques, ses sources subjectives*, Ann. Inst. Henri Poincaré **7** (1937), no. 1, 1–68.

[31] R. Pettigrew, *Dutch Book Arguments, Series: Elements in Decision Theory and Philosophy*, Cambridge University Press, Cambridge, 2020.

[32] G. W. Leibniz, *On estimating the uncertain*, Leibniz Rev. **14** (2004), 43–53.

[33] L. Daston, *Classical Probability in the Enlightenment*, Princeton University Press, Princeton, 1988.

[34] C. Howson, *Can logic be combined with probability? Probably*, J. Appl. Logic **7** (2009), no. 2, 177–187.

[35] G. Coletti and R. Scozzafava, *Probabilistic Logic in a Coherent Setting*, Kluwer Academic Publishers, Dordrecht, 2002.

[36] T. Flaminio, L. Godo, and H. Hosni, *On the logical structure of de Finettis notion of event*, J. Appl. Logics **12** (2014), no. 3, 279–301.

[37] T. Flaminio, H. Hosni, and F. Montagna, *Strict coherence on many valued algebras*, J. Symb. Log. **83** (2018), no. 1, 55–69.

[38] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

[39] H. Gaifman, *Concerning measures on Boolean algebras*, Pacific J. Math. **14** (1964), no. 1, 61–73.

[40] A. Horn and A. Tarski, *Measures in Boolean algebras*, Trans. Amer. Math. Soc. **64** (1948), 467–497.

[41] T. Jech, *Measures on Boolean algebras*, Fund. Math. **239** (2017), no. 2, 177–183.

[42] H. Gaifman and M. Snir, *Probabilities over rich languages, testing and randomness*, J. Symb. Log. **47** (1982), no. 3, 495–548.

[43] S. L. Zabell, *Symmetry and its Discontents*, Cambridge University Press, Cambridge, 2011.

[44] J. Y. Halpern and D. Koller, *Representation dependence in probabilistic inference*, J. Artificial Intelligence Res. **21** (2004), 319–356.

[45] J. W. Romeyn, *Hypotheses and inductive predictions*, Synthese **141** (2004), no. 3, 333–364.

[46] J. B. Paris, *What you see is what you get*, Entropy **16** (2014), no. 11, 6186–6194.

[47] S. Wenmackers and J.-W. Romeijn, *New theory about old evidence*, Synthese **193** (2016), no. 4, 1225–1250.

[48] C. E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. **27** (1948), 379–423.

[49] C. E. Shannon, *Prediction and entropy of printed English*, Bell Syst. Tech. **30** (1951), no. 1, 50–64.

[50] J. Kuhr and D. Mundici, *De Finetti theorem and Borel states in* [0,1]*-valued algebraic logic*, Internat. J. Approx. Reason. **46** (2007), no. 3, 605–616.

[51] D. Mundici, *Bookmaking over infinite-valued events*, Internat. J. Approx. Reason. **43** (2006), no. 3, 223–240.

[52] R. L. O. Cignoli, I. M. L. D'Ottaviano, and D. Mundici, *Algebraic Foundations of Many-Valued Reasoning*, Kluwer Academic Publishers, Dordrecht, 1999.

[53] A. Hájek and N. Hall, *The hypothesis of the conditional construal of conditional probability*, in: E. Ells and B. Skyrms (Eds.), Probability and Conditionals: Belief Revision and Rational Decision, Cambridge University Press, Cambridge, 1994, pp. 75–111.

[54] D. Mundici, *Advanced Łukasiewicz Calculus and MV-algebras*, Springer, Dordrecht Heidelberg, London, New York, 2011.

[55] P. Baldi and H. Hosni, *Probability and degrees of truth*, The Logica Yearbook, College Publications, London, 2022, pp. 1–18.

[56] T. Flaminio, *Three characterizations of strict coherence on infinite-valued events*, Rev. Symb. Log. **13** (2020), no. 3, 593–610.

[57] P. Cintula, P. Hajek, and C. Noguera, *Handbook of Mathematical Fuzzy Logic*, Vol. 1–3, College Publications, London, 2011.

[58] P. Hajek, L. Godo, and F. Esteva, *Fuzzy logic and probability*, in: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI1995), 1995, pp. 237–244.

[59] P. Hájek, *Metamathematics of Fuzzy Logic*, Kluwer Academic Publishers, Dordrecht, 1998.

[60] P. Baldi, P. Cintula, and C. Noguera, *Classical and fuzzy two-layered modal logics for uncertainty: Translations and proof-theory*, Int. J. Comput. Int. Sys. **13** (2020), no. 1, 988–1001.

[61] R. Fagin, J. Y. Halpern, and N. Megiddo, *A logic for reasoning about probabilities*, Inform. Comput. **87** (1990), no. 1–2, 78–128.

[62] J.-Y. Jaffray, *Coherent bets under partially resolving uncertainty and belief functions*, Theory Decision **26** (1989), no. 2, 99–105.

[63] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[64] T. Flaminio, L. Godo, and H. Hosni, *Coherence in the aggregate: A betting method for belief functions on many-valued events*, Int. J. Approx. Reason. **58** (2015), 71–86.

[65] J. M. Keynes, *The general theory of employment*, Q. J. Econ. **51** (1937), no. 2, 209–223.

[66] D. Ellsberg, *Risk, ambiguity, and the savage axioms*, Q. J. Econ. **75** (1961), no. 4, 643–669.

[67] G. Shafer, *Perspectives on the theory and practice of belief functions*, Internat. J. Approx. Reason. **4** (1990), no. 5–6, 323–362.

[68] M. C. M. Troffaes and G. de Cooman, *Lower Previsions*, Wiley, Chichester, 2014.

[69] P. Baldi and H. Hosni, *Depth-bounded Belief functions*, Int. J. Approx. Reason. **123** (2020), 26–40.

[70] P. Baldi and H. Hosni, *A logic-based tractable approximation of probability*, J. Logic Comput. **33** (2023), no. 3, 599–622.

[71] E. A. Corsi, T. Flaminio, and H. Hosni, *Scoring rules for Belief functions and imprecise probabilities: A comparison*, in: J. Vejnarová, N. Wilson (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2021, Lecture Notes in Computer Science, Vol. 12897, Springer, Cham, 2021, pp. 301–313.

[72] E. A. Corsi, T. Flaminio, and H. Hosni, *When Belief functions and lower probabilities are indistinguishable*, in: A. Cano, J. De Bock, E. Miranda, S. Moral (Eds.), Proceedings of Machine Learning Research ISIPTA 2021, PMLR, vol. 147, 2021, pp. 83–89.

[73] R. Carnap, *The two concepts of probability: The problem of probability*, Philos. Phenomen. Res. **5** (1945), no. 4, 513–532.

[74] R. Carnap, *Logical Foundations of Probability*, 2nd ed., University of Chicago Press, Chicago, 1962.

[75] W. E. Johnson, *Probability: The deductive and inductive problems*, Mind **41** (1932), no. 164, 409–423.

[76] R. Carnap, *On inductive logic*, Philos. Sci. **12** (1945), no. 2, 72–97.

[77] J. B. Paris and A. Vencovská, *Pure Inductive Logic*, Cambridge University Press, Cambridge, 2015.

[78] Š. Dautović, D. Doder, and Z. Ognjanović, *Logics for reasoning about degrees of confirmation*, J. Logic Comput. **31** (2021), no. 8, 2189–2217.

[79] T. Groves, *Lakatosas criticism of Carnapian inductive logic was mistaken*, J. Appl. Logic **14** (2016), 3–21.

[80] H. Leitgeb, *Logic in general philosophy of science: old things and new things*, Synthese **179** (2011), no. 2, 339–350.

[81] H. Leitgeb and A. Carus, *Rudolf Carnap*, in: E. N. Zalta (Ed.), Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, Stanford, 2020.

[82] S. L. Zabell, *Carnap and the logic of inductive inference*, in: D. M. Gabbay, S. Hartmann, and J. Woods (Eds.), Handbook of the History of Logic, Elsevier, London, 2011, pp. 265–309.

[83] P. A. Schlipp, *The Philosophy of Rudolf Carnap*, Open Court, La Salle, 1963.

[84] J. Landes, J. B. Paris, and A. Vencovská, *Representation theorems for probability functions satisfying spectrum exchangeability in inductive logic*, Internat. J. Approx. Reason. **51** (2009), no. 1, 35–55.

[85] E. Howarth and J. B. Paris, *Pure inductive logic with functions*, J. Symb. Log. **84** (2019), 1382–1402.

[86] H. Gaifman, *Concerning measures in first order calculi*, Israel J. Math. **2** (1964), no. 1, 1–18.

[87] P. Diaconis and D. Freedman, *Partial exchangeability and sufficiency*, in: J. K. Ghosh and J. Roy (Eds.), Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, Indian Statistical Institute, 1984, pp. 205–236.

[88] A. Vencovská, *Binary induction and Carnap's continuum*, in: Proceedings of the Seventh Workshop on Uncertainty Processing (WUPES), Mikulov, 2006, pp. 173–182.

[89] J. Landes, *The Principle of Spectrum Exchangeability within Inductive Logic*, PhD thesis, Manchester Institute for Mathematical Sciences, 2009.

[90] A. Vencovská, *Extending Carnap's continuum to binary relations*, in: M. Banerjee and S. N. Krishna (Eds.), Logic and Its Applications, Lecture Notes in Computer Science (LNCS), Vol. 8923, Springer, 2015, pp. 207–217.

[91] J. F. C. Kingman, *Random partitions in population genetics*, Proc. R. Soc. Lon. Ser-A **361** (1978), no. 1704, 1–20.

[92] J. F. C. Kingman, *The representation of partition structures*, J. Lond. Math. Soc. **s2-18** (1978), no. 2, 374–380.

[93] H. Crane, *The ubiquitous Ewens sampling formula*, Statist. Sci. **31** (2016), no. 1, 1–19.

[94] J. B. Paris and A. Vencovská, *Six problems in pure inductive logic*, J. Philos. Logic **48** (2019), 731–747.

[95] J. B. Paris, *The Uncertain Reasoner's Companion: A Mathematical Perspective*, Cambridge Tracts in Theoretical Computer Science, Vol. 39, 2nd ed., Cambridge University Press, Cambridge, 2006.

[96] J. B. Paris and A. Vencovská, *A method for updating that justifies minimum cross entropy*, Internat. J. Approx. Reason. **7** (1992), no. 1–2, 1–18.

[97] J. B. Paris and A. Vencovská, *A note on the inevitability of maximum entropy*, Internat. J. Approx. Reason. **4** (1990), no. 3, 183–223.

[98] J. B. Paris and A. Vencovská, *In defense of the maximum entropy inference process*, Internat. J. Approx. Reason. **17** (1997), no. 1, 77–103.

[99] J. B. Paris and A. Vencovská, *Proof systems for probabilistic uncertain reasoning*, J. Symb. Log. **63** (1998), no. 3, 1007–1039.

[100] J. B. Paris and A. Vencovská, *Common sense and stochastic independence*, in: D. Corfield, J. Williamson (Eds.), Foundations of Bayesianism, Kluwer, Dordrecht, 2001, pp. 203–240.

[101] J. B. Paris, *Common sense and maximum entropy*, Synthese **117** (1998), no. 1, 75–93.

[102] J. Landes and G. Masterton, *Invariant equivocation*, Erkenntnis **82** (2017), 141–167.

[103] J. B. Paris and A. Vencovská, *On the applicability of maximum entropy to inexact reasoning*, Internat. J. Approx. Reason. **3** (1989), no. 1, 1–34.

[104] I. Csiszár, *Axiomatic characterizations of information measures*, Entropy **10** (2008), no. 3, 261–273.

[105] R. D. Rosenkrantz, *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*, Reidel, Dordrecht, 1977.

[106] L. J. Savage, *Elicitation of personal probabilities and expectations*, J. Amer. Statist. Assoc. **66** (1971), no. 336, 783–801.

[107] J. E. Shore and R. W. Johnson, *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*, IEEE T. Inform. Theory **26** (1980), no. 1, 26–37.

[108] M. Tribus, *Rational Descriptions, Decisions and Designs*, Pergamon Press, New York, 1969.

[109] H. Hosni and J. B. Paris, *Rationality as conformity*, Synthese **144** (2005), no. 2, 249–285.

[110] P. D. Grünwald and A. P. Dawid, *Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory*, Ann. Statist. **32** (2004), no. 4, 1367–1433.

[111] J. Landes and J. Williamson, *Objective Bayesianism and the maximum entropy principle*, Entropy **15** (2013), no. 9, 3528–3591.

[112] F. Topsøe, *Information theoretical optimization techniques*, Kybernetika **15** (1979), 1–27.

[113] V. Crupi, J. Nelson, B. Meder, G. Cevolani, and K. Tentori, *Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search*, Cognitive Sci. **42** (2018), 1410–1456.

[114] H. Cui, Q. Liu, J. Zhang, and B. Kang, *An improved Deng entropy and its application in pattern recognition*, IEEE Access **7** (2019), 18284–18292.

[115] R. Hanel, S. Thurner, and M. Gell-Mann, *Generalized entropies and the transformation group of superstatistics*, Proc. Nat. Acad. Sci. **108** (2011), no. 16, 6390–6394.

[116] J. Landes, *Probabilism, entropies and strictly proper scoring rules*, Int. J. Approx. Reason. **63** (2015), 1–21.

[117] G. Wilmers, *A foundational approach to generalising the maximum entropy inference process to the multi-agent context*, Entropy **17** (2015), no. 2, 594–645.

[118] C. Beierle, M. Finthammer, and G. Kern-Isberner, *Relational probabilistic conditionals and their instantiations under maximum entropy semantics for first-order knowledge bases*, Entropy **17** (2015), no. 2, 852–865.

[119] G. Kern-Isberner, *Characterizing the principle of minimum cross-entropy within a conditional-logical framework*, Artificial Intelligence **98** (1998), no. 1–2, 169–208.

[120] G. Kern-Isberner, *Conditionals in Nonmonotonic Reasoning and Belief Revision*, Springer, Berlin, 2001.

[121] J. B. Paris and S. R. Rad, *Inference processes for quantified predicate knowledge*, in: W. Hodges, R. de Queiroz, (Eds.), Proceedings of WoLLIC, Lecture Notes in Computer Science (LNCS), Vol. 5110, Springer, 2008, pp. 249–259.

[122] O. Barnett and J. B. Paris, *Maximum entropy inference with quantified knowledge*, Logic J. IGPL **16** (2008), no. 1, 85–98.

[123] S. Rafiee Rad, *Probabilistic characterisation of models of first-order theories*, Ann. Pure Appl. Logic **172** (2021), no. 1, 102875.

[124] J. B. Paris and S. R. Rad, *A note on the least informative model of a theory*, in: F. Ferreira, B. Löwe, E. Mayordomo, L. MendesGomes (Eds.), Proceedings of CiE, Springer, Berlin, 2010, pp. 342–351.

[125] J. Landes, *The entropy-limit (conjecture) for $\Sigma_2$-premises*, Studia Logica **109** (2021), 423–442.

[126] J. Landes, *A triple uniqueness of the maximum entropy approach*, in: J. Vejnarová, N. Wilson (Eds.), Proceedings of ECSQARU, Springer, Cham, 2021, pp. 644–656.

[127] J. Landes, S. R. Rad, and J. Williamson, *Determining maximal entropy functions for objective Bayesian inductive logic*, J. Philos. Logic **52** (2023), 555–608.

[128] J. Landes and J. Williamson, *Justifying objective Bayesianism on predicate languages*, Entropy **17** (2015), no. 4, 2459–2543.

[129] J. Williamson, *In Defence of Objective Bayesianism*, Oxford University Press, Oxford, 2010.

[130] J. Williamson, *Lectures on Inductive Logic*, Oxford University Press, Oxford, 2017.

[131] J. Landes, S. R. Rad, and J. Williamson, *Towards the entropy-limit conjecture*, Ann. Pure Appl. Logic **172** (2021), no. 2, 102870.

[132] J. McCarthy, *Epistemological problems of artificial intelligence*, in: Proceedings of IJCAI, Vol. 2, 1977, pp. 1038–1044.

[133] R. Haenni, J.-W. Romeijn, J. Williamson, and G. Wheeler, *Probabilistic Logics and Probabilistic Networks*, Springer, Dordrecht, 2011.

[134] A. Darwiche, *Three modern roles for logic in AI*, Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Association for Computing Machinery (ACM), New York, 2020, pp. 229–243.

[135] A. Darwiche, *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, Cambridge, 2009.

[136] R. E. Neapolitan, *Learning Bayesian Networks*, Pearson, Upper Saddle River, 2003.

[137] G. F. Cooper, *The computational complexity of probabilistic inference using Bayesian belief networks*, Artificial Intelligence **42** (1990), no. 2, 393–405.

[138] J. Landes and J. Williamson, *Objective Bayesian nets for integrating consistent datasets*, J. Artificial Intelligence Res. **74** (2022), 393–458.

[139] J. B. Paris, *On filling-in missing conditional probabilities in causal networks*, Internat. J. Uncertain. Fuzziness Knowledge-Based Syst. **13** (2005), no. 3, 263–280.

[140] J. Williamson, *Bayesian Nets Bayesian Nets and Causality*, Oxford University Press, Oxford, 2005.

[141] M. Drton and M. H. Maathuis, *Structure learning in graphical modeling*, Annu. Rev. Stat. Appl. **4** (2017), no. 1, 365–393.

[142] J. Pearl, *Causality Models, Reasoning and Inference,* 2nd ed., Cambridge University Press, Cambridge, 2009.

[143] D. Alrajeh, H. Chockler, and J. Y. Halpern, *Combining experts' causal judgments*, Artificial Intelligence **288** (2020), 103355.

[144] D. Danks, C. Glymour, and R. E. Tillman, *Integrating locally learned causal structures with overlapping variables*, in: D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.), Proceedings of NIPS, Curran Associates, 2008, pp. 1665–1672.

[145] S. Triantafillou, I. Tsamardinos, and I. Tollis, *Learning causal structure from overlapping variable sets*, J. Mach. Learn. Res. **9** (2010), 860–867.

[146] C. Mayo-Wilson, *The limits of piecemeal causal inference*, Brit. J. Philos. Sci. **65** (2014), no. 2, 213–249.

[147] J. Y. Halpern, *Actual Causality*, MIT Press, Cambridge, 2016.

[148] D. Draheim, *Generalized Jeffrey Conditionalization*, Springer, Cham, 2017.

[149] C. P. De Campos and F. G. Cozman, *The inferential complexity of Bayesian and credal networks*, in: Proceedings IJCAI, Morgan Kaufmann, San Francisco, 2005, pp. 1313–1318.

[150] T. Lukasiewicz, *Credal networks under maximum entropy*, in: Proceedings of UAI, Morgan Kaufmann, San Francisco, 2000, pp. 363–370.

[151] S. Benferhat, P. Leray, and K. Tabia, *Belief graphical models for uncertainty representation and reasoning*, in: P. Marquis, O. Papini, and H. Prade (Eds.), A Guided Tour of Artificial Intelligence Research, Springer, Cham, 2020, pp. 209–246.

[152] H. J. Keisler, *Probability quantifiers*, in: J. Barwise and S. Feferman (Eds.), Model Theoretic Logics, Perspectives in Logic, Cambridge University Press, Cambridge, 1985, pp. 509–556.

[153] Z. Marković, Z. Ognjanović, and M. Rašković, *A probabilistic extension of intuitionistic logic*, MLQ Math. Log. Q. **49** (2003), no. 4, 415–424.

[154] A. Hájek, *What conditional probability could not be*, Synthese **137** (2003), no. 3, 273–323.

[155] J. Y. Halpern, *An analysis of first-order logics of probability*, Artificial Intelligence **46** (1990), no. 3, 311–350.

[156] J. Y. Halpern and R. Pucella, *A logic for reasoning about evidence*, J. Artificial Intelligence Res. **26** (2006), 1–34.

[157] Z. Ognjanovic, M. Raskovic, and Z. Markovic, *Probability Logics – Probability based formalization of uncertainty reasoning*, Springer, Cham, 2016.

[158] K. R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, London, 1962.

[159] G. Ras, N. Xie, M. Van Gerven, and D. Doran, *Explainable deep learning: a field guide for the uninitiated*, J. Artificial Intelligence Res. **73** (2022), 329–397.

[160] C. S. Calude and G. Longo, *The Deluge of spurious correlations in big data*, Found. Sci. **22** (2017), no. 3, 595–612.

[161] H. Hosni and A. Vulpiani, *Forecasting in light of big data*, Philos. Tech. **31** (2018), 557–569.

[162] R. Fagin, *Probabilities on finite models*, J. Symb. Log. **41** (1976), no. 1, 50–58.

[163] S. Fajardo and H. J. Keisler, *Model Theory of Stochastic Processes*, 2nd ed., Cambridge University Press, Cambridge, 2016.

[164] J. V. Glebskii, D. I. Kogan, M. I. Liogon'kii, and V. A. Talanov, *Volume and fraction of satisfiability of formulas of the lower predicate calculus*, Kibernetika **2** (1969), 17–27.

[165] J. Rabold, G. Schwalbe, and U. Schmid, *Expressive explanations of DNNs by combining concept analysis with ILP*, in: U. Schmid, F. Klügl, D. Wolter (Eds.), Proceedings of KI, Springer, Cham, 2020, pp. 148–162.