Chapter 16

PROBABILITY, LOGIC, AND PROBABILITY LOGIC

Alan Hájek

In *The Blackwell Companion to Logic,* ed. Lou Goble, Blackwell, 362-384.

The true logic of the world is in the calculus of probabilities
James Clerk Maxwell

1. Probability and logic

"Probability logic" might seem like an oxymoron. Logic traditionally concerns matters immutable, necessary and certain, while probability concerns the uncertain, the random, the capricious. Yet our subject has a distinguished pedigree. Ramsey begins his classic "Truth and Probability" [44] with the words: "In this essay the Theory of Probability is taken as a branch of logic...". De Finetti [7] speaks of "the logic of the probable". And more recently, Jeffrey [25] regards probabilities as estimates of truth values, and thus probability theory as a natural outgrowth of two-valued logic—what he calls "probability logic". However we put the point, probability theory and logic are clearly intimately related. This chapter explores some of the multifarious connections between probability and logic, and focuses on various philosophical issues in the foundations of probability theory.

Our survey begins in §2 with the probability calculus, what Adams [1, p. 34] calls "pure probability logic". As we will see, there is a sense in which the axiomatization of probability presupposes deductive logic. Moreover, some authors see probability theory as the proper framework for *inductive* logic—a formal apparatus for codifying the degree of support that a piece of evidence lends a hypothesis, or the impact of evidence on rational opinion.

Fixing a meaning of ‘probability’ allows us to draw more specific connections to logic. Thus we will consider in §3 various interpretations of probability. According to the classical interpretation, probability and possibility are intimately related, so that probability becomes a kind of modality. For objective interpretations such as the frequency and propensity theories, probability theory can be regarded as providing the logic of ‘chance’. By the lights of the subjective (Bayesian) interpretation, probability can be thought of as the logic of partial belief. And for the logical interpretation, the connection to logic is the most direct, probability theory being a logic of partial entailment, and thus a true generalization of deductive logic.

Kolmogorov’s axiomatization is the orthodoxy, what might be thought of as the probabilistic analogue of classical logic. However, a number of authors offer rival systems, analogues of ‘deviant’ logics, as it were—we will discuss some of these in §4, noting some bridges between them and various logics. Probabilistic semantics is introduced in §5. The conclusions of even valid inferences can be uncertain when the premises of the inferences are themselves uncertain. This prompts Adams’ version of “probability logic”, the study of the propagation of probability in such inferences. This in turn motivates our discussion in §6 of the literature on probabilities of conditionals, in which probability theory is used to illuminate the logic of conditionals.

One cannot hope for a complete treatment of a topic this large in a survey this short. The reader who is interested in pursuing these themes further is invited to consult the bibliographical notes and the bibliography at the end.

2. The probability axioms

Probability theory was inspired by games of chance in 17th century France and inaugurated by the Fermat-Pascal correspondence. Their work culminated in the publication of *The Port Royal Logic*, which offered a "logic of uncertain expectation" in Jeffrey's phrase [25]. However, the development of the probability calculus had to wait until well into the 20th century.

Kolmogorov begins his classic book [28] with what he calls the "elementary theory of probability": the part of the theory that applies when there are only finitely many events in question. Let Ω be a set (the 'universal set'). A *field* (or *algebra*) on Ω is a set of subsets of Ω that has Ω as a member, and that is closed under complementation (with respect to Ω) and finite union. Let Ω be given, and let $\mathcal{F}$ be a field on Ω. Kolmogorov's axioms constrain the possible assignments of numbers, called *probabilities*, to the members of $\mathcal{F}$. Let P be a function from $\mathcal{F}$ to [0, 1] obeying:

1. (Non-negativity) $P(A) \geq 0$ for all $A \in \mathcal{F}$
2. (Normalization) $P(\Omega) = 1$
3. (Finite additivity) $P(A \cup B) = P(A) + P(B)$ for all $A, B \in \mathcal{F}$ such that $A \cap B = \varnothing$

Call such a triple of $(\Omega, \mathcal{F}, P)$ a *probability space*.

Here the arguments of the probability function are sets, probability theory being thus parasitic on set theory. We could instead attach probabilities to members of a collection S of *sentences* of a language, closed under finite truth-functional combinations, with the following counterpart axiomatization:

I. $P(A) \geq 0$ for all $A \in S$.

II. If T is a tautology (of classical logic), then $P(T) = 1$.

III. $P(A \vee B) = P(A) + P(B)$ for all $A \in S$ and $B \in S$ such that A and B are logically incompatible.

Note how these axioms take the notions of 'tautology' and 'logical incompatibility' as antecedently understood. To this extent we may regard probability theory as parasitic on deductive logic.

Kolmogorov then allows Ω to be infinite. A non-empty collection $\mathcal{F}$ of subsets of Ω is called a *sigma field* (or *sigma algebra*, or *Borel field*) on Ω iff $\mathcal{F}$ is closed under complementation and countable union. Define a *probability measure* P(–) *on* $\mathcal{F}$ as a function from $\mathcal{F}$ to [0, 1] satisfying axioms 1-3, as before, and also:

4. (Continuity) $E_n \downarrow \emptyset$ fi $P(E_n) \to 0$ (where $E_n \in \mathcal{F}\ \forall n$)

Equivalently, we can replace the conjunction of axioms 3 and 4 with a single axiom:

3′. (Countable additivity) If $\{A_i\}$ is a countable collection of (pairwise) disjoint sets, each $\in \mathcal{F}$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

The conditional probability of X given Y is standardly given by the ratio of unconditional probabilities:

$$P(X|Y) = \frac{P(X \leftrightarrow Y)}{P(Y)}\text{, provided } P(Y) > 0.$$

This is often taken to be the *definition* of conditional probability, although it should be emphasized that this is a technical usage of the term that may not align perfectly with a pretheoretical concept that we might have. For example, it seems that we can make sense of conditional probabilities that are defined in the absence of the requisite unconditional probabilities, or when the condition Y has probability 0 (Hájek [18]).

We can now prove various important theorems, among them:

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) \quad \text{(the law of total probability)}$$

Even more importantly, we can prove various versions of *Bayes' theorem*:

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

$$= \frac{P(B|A).P(A)}{P(B|A).P(A) + P(B|\neg A).P(\neg A)}$$

These are all of the essentials of the mathematical theory of probability that we will need here.[1] Jeffrey [25] stresses various analogies between this "formal probability logic" and deductive logic. Here is an important one: theorems such as these enable us, given certain probabilities, to calculate further probabilities. However, the probability calculus does not itself determine the probabilties of any sentences, apart from 1 for tautologies and 0 for contradictions. Such values need to be provided 'from the outside', with probability theory only providing a framework. Compare this with deductive logic, which tells us which sentences are consistent with others, and which sentences are implied by others, but which does not itself determine the truth values of any sentences, apart from 'true' for tautologies and 'false' for contradictions.

How, then, are probability values determined in the first place? This brings us to the issue of what probabilities *are*—that is, to the so-called *interpretation* of probability.

3. Interpretations of probability

The mathematics of probability is well understood, but its interpretation is controversial. Kolmogorov's axioms are remarkably economical, and as such they admit of many interpretations. We briefly present some of the best known ones, emphasizing various connections to logic along the way.[2]

3.1 The classical interpretation

According to the classical interpretation, championed by Laplace among others, in the face of evidence equally favoring each of various possibilities, or no such evidence at all, the probability of an event is simply the fraction of the total number of possibilities in which the event occurs—this is sometimes called the *principle of indifference*. Thus, the modalities of possibility and probability are intimately related. For example, the probability of a fair die landing with an even number showing up is 3/6. Unfortunately, the prescription can apparently yield contradictory results when there is no single privileged set of possibilities. And even when there is, critics have argued that biases cannot be ruled out *a priori*. Finally, classical probabilities are only finitely additive, so they do not provide an interpretation of the full Kolmogorov calculus.

3.2 The logical interpretation

Logical theories of probability are descendants of the classical theory. They generalize the notion that probability is to be computed in the absence of evidence, or on the basis of symmetrically balanced evidence, to allow probability to be computed on the basis of the evidence, whatever it may be. At least in their earlier forms, logical theories saw the probability of a hypothesis given such evidence as objectively and uniquely determined, and thus ideally to be agreed on by all rational agents. If we think that there can be, besides deductive implication, a weaker relation of partial implication, then we may also think of logical probability as an analysis of 'degree of implication'. This interpretation more than any other explicitly sees probability as part of logic, namely *inductive* logic.

Early proponents of logical probability include Keynes, W.E. Johnson, and Jeffreys. However, by far the most systematic study of logical probability was by Carnap. He thought of probability theory as an elaboration of deductive logic, arrived at by adding extra rules. Specifically, he sought to explicate 'the degree to which hypothesis h is confirmed by evidence e', with the 'correct' conditional probability c(h, e) its explication. Statements of logical probability such as 'c(h, e) = x' were then to be thought of as logical truths.

The formulation of logical probability begins with the construction of a formal language. Carnap [4] initially considers a class of very simple languages consisting of a finite number of logically independent monadic predicates (naming properties) applied to countably many individual constants (naming individuals) or variables, and the usual logical connectives. The strongest (consistent) statements that can be made in a given language describe all of the individuals in as much detail as the expressive power of the language allows. They are conjunctions of complete descriptions of each individual, each description itself a conjunction containing exactly one occurrence (negated or unnegated) of each predicate letter of the language. Call these strongest statements *state descriptions.*

An *inductive logic* for a language is a specification, for each pair of statements <p, q> of the language, a unique probability value, or degree of confirmation c(p, q). To achieve this, we begin by defining a probability measure m(–) over the state descriptions. Every sentence h of a given language is equivalent to a disjunction of (mutually exclusive) state descriptions, and its a priori probability m(h) is thus determined. m in turn will induce a confirmation function c(–,–) according to the conditional probability formula:

$$c(h, e) = \frac{m(h \,\&\, e)}{m(e)}$$

There are obviously infinitely many candidates for such an m, and hence c, even for very simple languages. However, Carnap favours one particular measure, which he calls “m*”. He argues that the only thing that significantly distinguishes individuals from one another is some qualitative difference, not just a difference in labeling. Define a *structure description* as a maximal set of state descriptions, each of which can be obtained from another by some permutation of the individual names. m* assigns numbers to the state descriptions as follows: first, every structure description is assigned an equal measure; then, each state description belonging to a given structure description is assigned an equal share of the measure assigned to the structure description. From this, we can then define

$$c^*(h, e) = \frac{m^*(h \,\&\, e)}{m^*(e)}$$

m* gives greater weight to homogenous state descriptions than to heterogeneous ones, thus ‘rewarding’ uniformity among the individuals in accordance with putatively reasonable inductive practice. It can be shown that c* allows inductive learning from experience. However, even insisting that an acceptable confirmation function must allow such learning, there are still infinitely many candidates; we do not have yet a reason to think that c* is the right choice. Carnap realizes that there is some arbitrariness here, but nevertheless regards c* as the proper function for inductive logic—he thinks it stands out for being simple and natural.

He later generalizes his confirmation function to a continuum of confirmation functions c_λ, each of which gives the weighted average (weighted according to a positive real number λ) of an a priori value of the probability in question, and that calculated in the light of evidence. Define a *family* of predicates to be a set of predicates such that, for each individual, exactly one member of the set applies. Carnap goes on to explore first-order languages containing a finite number of families of predicates. In his [5] he considers the special case of a language containing only one-place predicates. He lays down a host of axioms concerning the confirmation function c, including those induced by the probability calculus itself, various axioms of symmetry (for example, that c(h, e) remains unchanged under permutations of individuals, and of predicates of any family), and axioms that guarantee undogmatic inductive learning, and long-run convergence to relative frequencies. They imply that, for a family $\{P_n\}$, n = 1, ..., k , k > 2:

$$c_\lambda \text{ (individual } s + 1 \text{ is } P_j, s_j \text{ of the first } s \text{ individuals are } P_j) = \frac{s_j + \lambda/k}{s + \lambda},$$

where λ is a positive real number.

The higher the value of λ is, the less impact evidence has.

The problem remains: what is the correct setting of λ? And problems remain even once we have fixed the value of λ. It turns out that a universal statement in an infinite universe always receives zero confirmation, no matter what the (finite) evidence. Many find this counterintuitive, since laws of nature with infinitely many instances can apparently be confirmed. Hintikka [20] provides a system of confirmation that avoids this problem.

Recalling an objection to the classical interpretation, the various axioms of symmetry are hardly mere truths of logic. More seriously, we cannot impose further symmetry constraints that are seemingly just as plausible as Carnap's, on pain of inconsistency—see Fine [13, p. 202]. Moreover, Goodman's 'grue' paradox apparently teaches us that some symmetries are better than others, and that inductive logic must be sensitive to the meanings of predicates, strongly suggesting that a purely syntactic approach such as Carnap's is doomed. One could try to specify a canonical language, free of such monstrosities as 'grue', to which such syntactic rules might apply. But traditional logic finds no need for such a procedure, sharpening the suspicion that Carnap's program is *not* merely that of extending the boundaries of logic. Scott and Krauss [48] use model theory in their formulation of logical probability for richer and more realistic languages than Carnap's. Still, finding a canonical language seems to many to be a pipe dream, at least if we want to analyze the "logical probability" of any argument of real interest—either in science, or in everyday life.

3.3 Frequency interpretations

The guiding empiricist idea of frequency interpretations, which originated with Venn, is that an event's probability is the *relative frequency* of events of that type within a suitably chosen reference class. The probability that a given coin lands 'heads', for example, might be identified with the relative frequency of 'heads' outcomes in the class of all tosses of that coin. But there is an immediate problem: observed relative frequencies can apparently come apart from true probabilities, as when a fair coin that is tossed ten times happens to land heads every time. Von Mises [58] offers a more sophisticated formulation based on the notion of a *collective,* rendered precise by Church: a hypothetical infinite sequence of 'attributes' (possible outcomes) of a specified experiment, for which the limiting relative frequency of any attribute exists, and is the same in any recursively specified subsequence. The probability of an attribute A, relative to a collective ω, is then defined as the limiting relative frequency of A in ω. Limiting relative frequencies violate countable additivity, and the domain of definition of limiting relative frequency is not even a field. Thus it does not genuinely provide an interpretation of Kolmogorov's probability calculus.

As well as giving a (hypothetical) limiting relative frequency interpretation of probabilities of events, Reichenbach [45] gives an interpretation in terms of truth frequencies: the probability of truth of a statement of a certain type is the limiting relative frequency of statements of that type being true in a specified reference class of statements. Out of this definition he constructs a probability logic with a continuum of truth values, corresponding to the various possible probabilities.

A notorious problem for any version of frequentism is the so-called *problem of the single case:* we sometimes attribute non-trivial probabilities to results of experiments that occur only once, and that indeed may do so of necessity. Moreover, frequentist probabilities are always relativised to a reference class, which needs to be fixed in a way that does not appeal to probability; but most events belong to *many* 'natural' reference classes, which need not agree on the required relative frequency. In this sense, there may be no such thing as *the* probability of a given event—the infamous *reference class problem*. The move to hypothetical infinite sequences of trials creates its own problems: There is apparently no fact of the matter as to what such a hypothetical sequence would be, nor even what its limiting relative frequency for a given attribute would be, nor indeed whether that limit is even defined; and the limiting relative frequency can be changed to any value one wants by suitably permuting the order of trials. In any case, the empiricist intuition that facts about probabilities are simply facts about patterns in the actual phenomena has been jettisoned. Still more sophisticated accounts, frequentist in spirit, uphold this intuition—see, for instance, Lewis (1994). Such accounts sacrifice another intuition: that it is built into the very concept of 'chanciness' that fixing what actually happens does *not* fix the probabilities.

3.4 Propensity interpretations

Attempts to locate probabilities 'in the world' are also made by variants of the *propensity* interpretation, championed by such authors as Popper, Mellor and Giere. Probability is thought of as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run relative frequency of such an outcome. This view is explicitly intended to make sense of single-case probabilities, such as ‘the probability that this radium atom decays in 1500 years is 1/2’. According to Popper [43], a probability p of an outcome of a certain type is a propensity of a repeatable experiment to produce outcomes of that type with limiting relative frequency p. With its heavy reliance on limiting relative frequency, this position risks collapsing into von Mises-style frequentism. It seems moreover not to be a genuine interpretation of the probability calculus at all, for the same reasons that limiting relative frequentism is not. Giere [15], on the other hand, explicitly allows single-case propensities, with no mention of frequencies: probability is just a propensity of a repeatable experimental set-up to produce sequences of outcomes. This, however, creates the opposite problem to Popper’s: how, then, do we get the desired connection between probabilities and frequencies? Indeed, it is not clear why the assignments of such propensities should obey the probability calculus at all. For reasons such as these, propensity accounts have been criticized for being unacceptably vague.

3.5 The subjectivist interpretation (subjective Bayesianism)

Degrees of belief

Subjectivism is the doctrine that probabilities can be regarded as degrees of belief, sometimes called *credences*. It is often called ‘Bayesianism’ thanks to the important role that Bayes’ theorem typically plays in the subjectivist’s calculations of probabilities, although the theorem itself is neutral regarding interpretation. Unlike the logical interpretation (at least as Carnap originally conceived it), subjectivism allows that different agents with the very same evidence can rationally give different probabilities to the same hypothesis.

But what is a degree of belief? A standard analysis invokes betting behaviour: an agent's degree of belief in A is p iff she is prepared to pay up to p units for a bet that pays 1 unit if A, 0 if not A (de Finetti [7]). It is assumed that she is also prepared to sell that bet for p units. We thus have an operational definition of subjective probability, and indeed it inherits some of the difficulties of operationalism in general, and of behaviourism in particular. For example, the agent may have reason to misrepresent her true opinion. Moreover, as Ramsey [44] points out, the proposal of the bet may itself alter her state of opinion; and she might have an eagerness or reluctance to bet. These problems are avoided by identifying her degree of belief in a proposition with the betting price she regards as fair, whether or not she enters into such a bet—see Howson and Urbach [22]. Still, the fair price of a bet on X appears to measure not her probability that X will be the case, but rather her probability that X will be the case *and* that the prize will be paid, which may be rather less—for example, if X is unverifiable. Some think that this commits proponents of the betting interpretation to an underlying intuitionistic logic.

If we placed no restriction on who the agent is, we would not have an interpretation of the probability calculus at all, for there would be no guarantee that her degrees of belief would conform to it. Human agents sometimes violate the probability calculus in alarming ways ([26]), and indeed conforming all our degrees of belief to the probability calculus is surely an impossible standard. However, if we restrict our attention to ideally rational agents, the claim that degrees of belief are at least finitely additive probabilities becomes more plausible (much as deductive consistency of our beliefs might be an impossible standard but a reasonable ideal).

So-called *Dutch Book* arguments provide one important line of justification of this claim. A Dutch Book is a series of bets, each of which the agent regards as fair, but which collectively guarantee her loss. De Finetti [7] proves that if your degrees of belief are not finitely additive probabilities, then you are susceptible to a Dutch Book. Equally important, and often neglected, is Kemeny's [27] converse theorem: If your degrees of belief are finitely additive probabilities, then no Dutch Book can be made against you.

A related defence of the probability axioms comes from *utility theory*. Ramsey [44] derives both probabilities and utilities (desirabilities) from rational preferences. Specifically, given various assumptions about the richness of the preference space, and certain "consistency" assumptions, he shows how to define a real-valued utility function of the outcomes—in fact, various such functions will represent the agent's preferences. It turns out that ratios of utility-differences are invariant, the same whichever representative utility function we choose. This fact allows Ramsey to define degrees of belief as ratios of such differences, and to show that they are finitely additive probabilities. However, it is dubious that *consistency* requires one to have a set of preferences as rich as Ramsey requires. This places strain on Ramsey's claim to assimilate probability theory to logic. However, Howson (1997) argues that a betting interpretation of probability underpins a soundness and completeness proof of the probability axioms, thus supporting the claim that the Bayesian theory does provide a logic of consistent belief.

Savage [47] likewise derives probabilities and utilities from preferences among options that are constrained by certain putative 'consistency' principles. Jeffrey [24] refines the method further, giving a "logic of decision" according to which rational choice maximizes *expected utility*, a certain probability-weighted average of utilities.

Updating Probability

Suppose that your degrees of belief are initially represented by a probability function $P_{initial}(–)$, and that you become certain of a piece of evidence E. What should be your new probability function P_{new}? Since you want to avoid any gratuitous changes in your degrees of belief that were not prompted by the evidence, P_{new} should be the minimal revision of $P_{initial}$ subject to the constraint that $P_{new}(E) = 1$. The favoured updating rule among Bayesians is *conditionalization:* P_{new} is derived from $P_{initial}$ by taking probabilities conditional on E, according to the schema:

(Conditionalization) $P_{new}(X) = P_{initial}(X|E)$ (provided $P_{initial}(E) > 0$)

Lewis [33] gives a 'diachronic' Dutch Book argument for conditionalization: if your updating is rule-governed, you are subject to a Dutch Book (at the hands of a bookie who knows the rule you employ) if you do not conditionalize. Equally important is the converse theorem: if you do conditionalize, you cannot be Dutch Booked (Skyrms [53]).

Now suppose that as the result of some experience your degrees of belief across a countable partition $\{E_1, E_2, ...\}$ change to $\{P_{new}(E_1), P_{new}(E_2), ..., \}$, where none of these values need be 1 or 0. The rule of *Jeffrey conditionalization*, or *probability kinematics,* relates your new probability function to the initial one according to:

$$P_{new}(X) = \sum_i P_{initial}(X|E_i)P_{new}(E_i)$$

(If the probabilities are only finitely additive, then the partition and sum must be finite.) Conditionalization can be thought of as the special case of Jeffrey conditionalization in which $P(E_i) = 1$ for some i. Jeffrey conditionalization is supported by a Dutch Book argument due to Armendt; it is also the rule that, subject to the constraints on the partition, minimizes a measure of 'distance' in function space between the initial and new probability functions, called 'cross-entropy' (Diaconis and Zabell [9]).

Orthodox Bayesianism can now be characterized by the following maxims:

B1) The rational agent's 'prior' (initial) probabilities conform to the probability calculus.

B2) The rational agent's probabilities update by the rule of (Jeffrey) conditionalization.

B3) There are no further constraints on the rational agent.

Some critics reject orthodox Bayesianism's radical permissiveness regarding prior probabilities. A standard defense (e.g., Savage [47], Howson and Urbach [22]) appeals to famous ‘convergence-to-truth’, and ‘merger-of-opinion’ results. Roughly, their content is that with probability one, in the long run the effect of choosing one prior rather than another is washed out. Successive conditionalizations on the evidence will make a given agent eventually converge to the truth, and thus, initially discrepant agents eventually come to agree with each other (assuming that the priors do not give probability zero to the truth, and that the stream of incoming evidence is sufficiently rich). In an important sense, at least this much inductive logic is implicit in the probability calculus.

But Bayesianism is a theme that admits of many variations.

Further constraints on subjective probabilities

Against B3), some less permissive Bayesians also require that a rational agent’s probabilities be *regular,* or *strictly coherent:* if $P(X) = 1$, then X is a tautology. Regularity is the converse of axiom II, again linking probability and logic. It is meant to guard against the sort of dogmatism that no course of learning by (Jeffrey) conditionalization could cure.

Van Fraassen (1995) suggests a further constraint on rational opinion called *Reflection*, involving credences about one’s own future credences. Here is one formulation:

$$\text{(Reflection)}\ P_t(A|P_{t+\Delta}(A) = x) = x \quad (\Delta > 0)$$

where P_t is the agent’s probability function at time t. The idea is that when all is well, a certain sort of epistemic integrity requires one to regard one’s future opinions as being trustworthy, having arisen as the result of a rational process of learning. A more general version of Reflection is presented by Goldstein [16].

Lewis [36] offers a principle that links objective chance and rational credence:

$$\text{(Principal Principle)} \quad P_t(A|ch_t(A) = x \ \&\ E) = x.$$

where $ch_t(A)$ is the objective chance of A at time t, and E is any information that is “admissible” at time t (roughly, gives no evidence about the actual truth value of A). For example, the principle tells us: given that I believe this coin is fair, and thus has a chance of 1/2 at the moment of landing heads at the next toss, I should assign credence 1/2 to that outcome. The principle can, on the other hand, be thought of as giving an implicit characterization of chance as a theoretical property whose distinctive role is to constrain rational credences in just this way. (See the articles by Lewis, Hall, and Thau in *Mind* Vol. 103 (1994) for refinements of the principle.)

Finally, there have been various proposals for resuscitating symmetry constraints on priors, in the spirit of the classical and logical interpretations. More sophisticated versions of the principle of indifference have been explored by Jaynes [23] and Paris and Venkovská [39]. Their guiding idea is to maximize the probability function's *entropy*, which for an assignment of positive probabilities $p_1, ..., p_n$ to n worlds equals $-\Sigma_i p_i \log(p_i)$.

Orthodox Bayesianism has also been denounced for being overly demanding: its requirements of sharp probability assignments to all propositions, logical omniscience, and so on have been regarded by some as unreasonable idealizations. This motivates various relaxations of tenets B1) and B2) above. B2) might be weakened to allow other rules for the updating of probabilities besides conditionalization—for example, revision to the probability function that maximizes entropy, subject to the relevant constraints (Jaynes [23], Skyrms [52]). And some Bayesians drop the requirement that probability updating be rule-governed altogether—see Earman [10].

The relaxation of B1) is a large topic, and it motivates some of the non-Kolmogorovian theories of probability, to which we now turn.

4. Non-Kolmogorovian theories of probability

A number of authors would abandon the search for an adequate interpretation of Kolmogorov's probability calculus, since they abandon some part of his axiomatization.

Abandoning the sigma field sub-structure

Fine [13] argues that requiring the domain of the probability function to be a sigma field is overly restrictive. For example, I might have limited census data on race and gender that give me good information concerning the probability P(M) that a randomly chosen person is male, and the probability P(B) that such a person is black, without giving me any information about the probability $P(M \cap B)$ that such a person is both male and black.

Abandoning sharp probabilities

Each Kolmogorovian probability is a single number. But suppose that your state of opinion does not determine a single probability function, but rather is consistent with a multiplicity of such functions. In that case, we might represent your opinion as the *set* of all these functions (e.g., Levi [32]; Jeffrey [25]). Each function in this set corresponds to a way of precisifying your opinion in a legitimate way. This approach will typically coincide with interval-valued probability assignments, but it need not. Koopman (in [30], 119-131) offers axioms for 'upper' and 'lower' probabilities which may be thought of as the endpoints of such intervals. See also Walley [60] for an extensive treatment of imprecise probabilities.

Abandoning numerical probabilities altogether

In contrast to the "quantitative" probabilities so far assumed, Fine [13] sympathetically canvases various theories of comparative probability, exemplified by statements of the form 'A is at least as probable as B' ($A \Rightarrow B$) He offers axioms governing '$\Rightarrow$', and explores the conditions under which comparative probability can be given a representation in terms of Kolmogorovian probabilities.

Negative and complex-valued probabilities

More radically, physicists such as Dirac, Wigner, and Feynman have countenanced *negative* probabilities. Feynman, for instance, suggests that a particle diffusing in one dimension in a rod has a probability of being at a given position and time that is given by a quantity that takes negative values. Depending on how we interpret probability, however, we may instead want to say that this function bears certain analogies to a probability function, but when it goes negative the analogy breaks down. Cox allows probabilities to take values among the complex numbers in his theory of stochastic processes having discrete states in continuous time. See Mückenheim [38] for references.

Abandoning the normalization axiom

It might seem to be entirely conventional that the maximal value a probability function can take is 1. However, it has some non-trivial consequences. Coupled with the other axioms, it guarantees that a probability function takes at least two distinct values, whereas setting $P(\Omega) = 0$ would not. More significantly, it is non-trivial that there *is* a maximal value. Other measures, such as length or volume, are not so bounded. Indeed, Renyi [46] drops the normalization assumption altogether, allowing probabilities to attain the 'value' ∞.

Some authors want to loosen the rein that classical logic has on probability, allowing logical/necessary truths to be assigned probability less than one—perhaps to account for the fact that logical or mathematical conjectures may be more or less well confirmed. See, for example, Polya [41]. Furthermore, axiom II makes reference to the notion of 'tautology', with classical logic implicitly assumed. Proponents of non-classical logics may wish to employ instead their favorite 'deviant' notion of 'tautology' (perhaps requiring corresponding adjustments elsewhere in the axiomatization). For example, *constructivist* theories grounds probability theory in intuitionistic logic.

Infinitesimal probabilities

Kolmogorov’s probability functions are real-valued. A number of philosophers (e.g., Lewis [36], Skyrms [50]) drop this assumption, allowing probabilities to take values from the real numbers of a *nonstandard model* of analysis—see Skyrms ([50, appendix 4]) for the construction of such a model. In particular, they allow probabilities to be *infinitesimal:* positive, but smaller than every (standard) real number. Various non-empty propositions in infinite probability spaces that would ordinarily receive probability zero according to standard probability theory, and thus essentially treated as impossible, may now be assigned positive probability. (Consider selecting at random a point from the [0, 1] interval.) In uncountable spaces, regular probability functions cannot avoid taking infinitesimal values.

Abandoning countable additivity

Kolmogorov’s most controversial axiom is undoubtedly *continuity*—that is, the ‘infinite part’ of countable additivity. He regarded it as an idealization that finessed the mathematics, but that had no empirical meaning. As we have seen, according to the classical, frequency, and certain propensity interpretations, probabilities violate countable additivity. De Finetti [8] marshals a battery of arguments against it. Here is a representative one: Countable additivity requires one to assign an extremely biased distribution to a denumerable partition of events. Indeed, for any $\varepsilon > 0$, however small, there will be a finite number of events that have a combined probability of at least $1 - \varepsilon$, and thus the lion’s share of all the probability.

Abandoning finite additivity

Various theories of probability that give up even finite additivity have been proposed—so-called *non-additive* probability theories.

Dempster-Shafer theory begins with a *frame of discernment* Ω, a partition of hypotheses. To each subset of Ω we assign a ‘mass’ between 0 and 1 inclusive; all the masses sum to 1. We define a *belief function* Bel(A) by the rule: for each subset A of Ω, Bel(A) is the sum of the masses of the subsets of A. Shafer [49] gives this interpretation: Suppose that the agent will find out for certain some proposition on Ω. Then Bel(A) is his degree of belief that he will find out A. Bel(A) + Bel(–A) need not equal 1; indeed, Bel(A) and Bel(–A) are functionally independent of each other. Belief functions have many of the same formal properties as Koopman's lower probabilities. Mongin [37] shows that there are important links between epistemic modal logics and Dempster-Shafer theory.

So-called “Baconian probabilities” represent another non-additive departure from the probability calculus. The Baconian probability of a conjunction is equal to the minimum of the probabilities of the conjuncts. Such ‘probabilities’ are formally similar to membership functions in fuzzy logic. L.J. Cohen [6] regards them as appropriate for measuring inductive support, and for assessing evidence in a court of law.

For further non-additive probability theories, see (among others) Ghirardato’s modeling of ambiguity aversion, Shackle’s potential surprise functions, Dubois and Prade’s theory of fuzzy probabilities, Schmeidler’s and Wakker’s respective theories of expected utility, and Spohn’s theory of non-probabilistic belief functions—Ghirardato [14] and Howson [21] have references and more discussion.

Conditional probability as primitive

According to each of the interpretations of probability that we have discussed, probability statements are always at least tacitly relativised. On the classical interpretation, they are relativised to the set of possibilities under consideration; on the logical interpretation, to an evidence statement; on the frequency interpretations, to a reference class; on the propensity interpretation, to a chance set-up; on the subjective interpretation, to a subject (who may have certain background knowledge) at a time. Perhaps, then, it is *conditional* probability that is the more fundamental notion.

Rather than axiomatizing unconditional probability, and later defining conditional probability therefrom, Popper [42] and Renyi [46] take conditional probability as primitive, and axiomatize it directly. Popper's system is more familiar to philosophers. His primitives are: (i) Ω, the universal set; (ii) a binary numerical function p(–,–) of the elements of Ω; a binary operation ab defined for each pair (a, b) of elements of Ω; a unary operation ¬a defined for each element a of Ω. Each of these concepts is introduced by a postulate (although the first actually plays no role in his theory):

Postulate 1. The number of elements in Ω is countable.

Postulate 2. If a and b are in Ω, then p(a, b) is a real number, and we have as axioms:

A1. (Existence) There are elements c and d in Ω such that

$$p(a, b) \neq p(c, d).$$

A2. (Substitutivity) If p(a, c) = p(b, c) for all c in Ω, then

$$p(d, a) = p(d, b) \text{ for all d in } \Omega.$$

A3. (Reflexivity) p(a, a) = p(b, b).

Postulate 3. If a and b are in Ω, then ab is in Ω; and if c is also in Ω, then we have:

B2. (Monotony) $p(ab, c) \leq p(a, c)$

B2. (Multiplication) $p(ab, c) = p(a, bc)p(b, c)$

Postulate 4. If a is in Ω, then ¬a is in Ω; and if b is also in Ω, then we have:

C. (Complementation) $p(a, b) + p(\neg a, b) = p(b, b)$, unless $p(b, b) = p(c, b)$ for all c in Ω.

Popper also adds a 'fifth postulate', which may be thought of as giving the definition of absolute (unconditional) probability:

Postulate AP. If a and b are in Ω, and if $p(b, c) \geq p(c, b)$ for all c in Ω, then

$$p(a) = p(a, b).$$

Here, b can be thought of as a tautology. Unconditional probability, then, is probability conditional on a tautology. Thus, Popper's axiomatization generalizes ordinary probability theory. A function p(–,–) that satisfies the above axioms is called a *Popper function*.

An advantage of using Popper functions is that conditional probabilities of the form p(a, b) can be defined, and can have intuitively correct values, even when b has absolute probability 0, rendering the usual conditional probability ratio formula inapplicable. For example, the probability that a randomly selected point from [0, 1] is 1/3, given E = 'it is either 1/3 or 2/3', is plausibly equal to 1/2, and a Popper function can yield this result; yet the probability of E is standardly taken to be 0. Popper functions also allow a natural generalization of updating by conditionalization, so that even items of evidence that were originally assigned probability 0 by an agent can be learned. McGee (in [11]) shows that, in an important sense, probability statements cast in terms of Popper functions and those cast in terms of nonstandard probability functions are inter-translatable.

5. Probabilistic semantics and probability propagation

5.1 Probabilistic semantics

Various notions from standard semantics can be recovered by *probabilistic semantics*. The central idea is to define the logical concepts in terms of probabilistic ones. Alternative axiomatizations of probability will then give rise to alternative logics.

Call a statement A of a first-order language L *logically true in the probabilistic sense* if for all probability functions P, P(A) = 1. Where S is a set of statements of L, say that A is *logically entailed by S in the probabilistic sense* if, for all P, P(A) = 1 if P(B) = 1 for each member B of S. This sense of logical entailment is strongly sound and strongly complete: S fi A iff A is logically entailed by S in the probabilistic sense. And taking S to be ∅, we have fi A iff A is logically true in the probabilistic sense. Popper functions also permit natural definitions of logical truth and logical entailment, with analogous soundness and completeness results—see Leblanc [31, pp. 245 ff.]. Van Fraassen [56] exploits (slightly differently axiomatized) primitive conditional probability functions in providing probabilistic semantics for intuitionistic propositional logic and classical quantifier logic. Probabilistic semantics have been supplied for first-order logic without and with identity, modal logic, and conditional logic—see Leblanc [31] for a good general survey and for references. Van Fraassen [57] offers such semantics for relevant logic; Pearl [40] has a general discussion of probabilistic semantics for nonmonotonic logic.

Probabilistic semantics represent a limiting case of the idea that a valid argument is one in which it is not possible for the probabilities of all of the premises to be high, while the probability of the conclusion is not. More generally, what can be said about the propagation of probability from the premises to the conclusion of a valid argument?

5.2 Probability propagation: Adams' probability logic

If the premises of a valid argument are all certain, then so is the conclusion. Suppose, on the other hand, that the premises are not all certain, but probable to various degrees; can we then put bounds on the probability of the conclusion? Or suppose that we want the probability of the conclusion of a given valid argument to be above a particular threshold; how probable, then, must the premises be? These questions are pressing, since in real-life arguments we typically are not certain of our premises, and it may be important to know how confidently we may hold their conclusions. Indeed, we know from the lottery paradox that each premise in a valid argument can be almost certain, while the conclusion is certainly false. 'Probability logic' is the name that Adams [2] gives to the formal study of such questions—the study of the transmission (or lack thereof) of probability through valid inferences. We sketch his treatment here.

The hallmark of his probability logic is that traditional concerns with truth and falsehood of premises and conclusions are replaced with concerns about their probabilities. This in turn leads to the non-monotonic nature of probability logic: a conclusion that is initially assigned high probability and hence accepted may later be retracted in the face of new evidence. Define the *uncertainty* u(F) of a sentence by

$$u(F) = 1 - P(F).$$

Various important results in probability logic are more conveniently stated in terms of uncertainties rather than probabilities. For example:

> *Valid inference uncertainty theorem (VIUT):* The uncertainty of the conclusion of a valid inference cannot exceed the sum of the uncertainties of the premises.

Hence, the uncertainty of the conclusion of a valid inference can only be large if the sum of the uncertainties of the premises is large—witness the lottery paradox, in which many small uncertainties in the premises accumulate to yield a maximally uncertain conclusion. In particular, if each premise has an uncertainty no greater than ε, then there must be at least $1/\varepsilon$ of them for the conclusion to have maximal uncertainty.

The VIUT gives a bound on the uncertainty of the conclusion. Under certain circumstances, the bound can be achieved. Call a premise of a valid inference *essential* if the inference that omits that premise but that is otherwise the same is invalid. We have:

> *Uncertainty bound attainment theorem:* Suppose $F_1, ..., F_n \therefore F$ is valid, and let $u_1, ..., u_n$ be nonnegative, with $\Sigma u_i \leq 1$. If the premises are consistent and all essential, then there is an uncertainty function u(–) such that $u(F_i) = u_i$ for i = 1, ..., n, and $u(F) = u_1 + ... + u_n$.

However, such 'worst case' uncertainties in a conclusion can be reduced by introducing redundancy among the premises from which it is derived. For it can further be shown that, given a valid inference with various premises, different subsets of which entail the conclusion, the conclusion's uncertainty cannot be greater than the total uncertainty of that subset with the smallest total uncertainty. Define a *minimal essential premise set* to be an essential premise set that has no proper subsets that are essential. Suppose that we have a valid inference with premises $F_1, ..., F_n$. The *degree of essentialness* of premise F_i, $e(F_i)$, is: *1/k*, where *k* is the cardinality of the smallest essential set of premises to which F_i belongs, if F_i belongs to some minimal essential set, and *0* otherwise. Intuitively, $e(F_i)$ is a measure of how much 'work' F_i does in a valid inference.

We can now generalize the VIUT:

Theorem: If $F_1, .., F_n \therefore F$ is valid, then $u(F) \leq e(F_1)u(F_1) + ... + e(F_n)u(F_n)$.

We thus can lower the upper bound on u(F) from that given by the VIUT.

Adams calls an inference *probabilistically valid* iff, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that, under any probability assignment according to which each of the premises has probability greater than $1 - \delta$, the conclusion has probability at least $1 - \varepsilon$. A linchpin of his account of probabilistic validity is his treatment of conditionals. According to him, a conditional has no truth value, and hence sense cannot be made of the probability of its *truth.* Yet conditionals can clearly figure either as premises or conclusions of arguments, and we still want to be able to assess these arguments. How, then, do we determine the probabilities of conditionals? This brings us to another important point of cross-fertilization between probability and logic.

6. Probabilities of conditionals

Probability and logic are intimately intertwined in the study of probabilities of conditionals. In the endeavor to furnish a logical analysis of natural language, the conditional has proved to be somewhat recalcitrant, and the subject of considerable controversy. [See Chapter 17] Meanwhile, the notion of 'conditionality' is seemingly well understood in probability theory, and taken by most to be enshrined in the usual ratio formula for conditional probability. Thus, fecund research programs have been founded on both promoting and parrying a certain marriage between the logic of conditionals and probability theory: the hypothesis that *probabilities of conditionals* are *conditional probabilities*. More precisely, the hypothesis is that some suitably quantified and qualified version of the following equation holds:

(PCCP) $P(A \rightarrow B) = P(B|A)$ for all A, B in the domain of P, with $P(A) > 0$.

('→' is a conditional connective.)

The best known presentations of this hypothesis are due to Stalnaker [54] and Adams [1]. Stalnaker hoped that a suitable version of it would serve as a criterion of adequacy for a truth-conditional account of the conditional. He explored the conditions under which it would be reasonable for a rational agent, with subjective probability function P, to believe a conditional A→B. By identifying the probability of A→B with P(B|A), Stalnaker was able to put constraints on the truth conditions of '→' (for example, the upholding of conditional excluded middle) that supported his preferred C2 logic. While Adams eschewed any truth-conditional account of the conditional, he was happy to speak of the probability of a conditional, equating it to the corresponding conditional probability. This allowed him to extend his notion of probabilistic validity to arguments that contain conditionals—arguments that on his view lie outside the scope of the traditional account of validity couched in terms of truth values. He argued that the resulting scheme respects intuitions about which inferences are reasonable, and which not.

With these motivations in mind, and for their independent interest, we consider four salient ways of rendering precise the hypothesis that probabilities of conditionals are conditional probabilities:

Universal version: There is some → such that for all P, (PCCP) holds.

Rational Probability Function version: There is some → such that for all rational subjective probability functions P, (PCCP) holds.

Universal Tailoring version: For each P there is some → such that (PCCP) holds.

Rational Probability Function Tailoring version: For each rational subjective probability function P, there is some → such that (PCCP) holds.

If any of these versions can be sustained, then important links between logic and probability theory will have been established, just as Stalnaker and Adams hoped. Probability theory would be a source of insight into the formal structure of conditionals; and probability theory in turn would be enriched, since we could characterize more fully what the usual conditional probability ratio means, and what its use is.

There is now a host of results—mostly negative—concerning (PCCP). Some preliminary definitions will assist in stating some of the most important ones. If (PCCP) holds (for a given → and P) we will say that → is a *PCCP-conditional for* P, and that P is a *PCCP-function for* →. If (PCCP) holds for each member P of a class of probability functions P, we will say that → is a *PCCP-conditional for* P. A pair of probability functions P and P' are *orthogonal* if, for some A, P(A) = 1 but P'(A) = 0. Call a proposition A a *P-atom* iff P(A) > 0 and, for all X, either P(AX) = P(A) or P(AX) = 0. Finally, we will call a probability function *trivial* if it has at most 4 different values.

The negative results are 'triviality results': only trivial probability functions can sustain PCCP, given certain assumptions, The earliest and most famous results are due to Lewis [33], which he later strengthens [35]. Their upshot is that there is no PCCP-conditional for any class of probability functions closed under conditionalizing (restricted to the propositions in a single finite partition), or under Jeffrey conditionalizing, unless the class consists entirely of trivial functions. These results refute the Universal version of the hypothesis. They also spell bad news for the Rational Probability Function version, since rationality surely permits having a non-trivial probability function and updating by (Jeffrey) conditionalizing. This version receives its death blow from a result by Hall (in [11]) that significantly strengthens Lewis' results:

Orthogonality result: Any two non-trivial PCCP-functions for a given → with the same domain are orthogonal.

It follows from this that the Rational Probability Function version is true only if any two rational agents' probability functions are orthogonal if distinct—which is absurd.

So far, the 'tailoring' versions remain unscathed. The Universal Tailoring version is refuted by the following result due to Hájek ([17], here slightly strengthened):

Finite probability functions result: Any non-trivial probability function with finite range has no PCCP-conditional.

This result also casts serious doubt on the Rational Probability Tailoring version, for it is hard to see why rationality requires one to have a probability function with infinite range. If we make a minimal assumption about the logic of the $\rightarrow$, matters are still worse thanks to another result of Hall's (1994):

No Atoms Result: Given (Ω, $\mathcal{F}$, P), suppose that PCCP holds for P and a '$\rightarrow$' that obeys modus ponens. Then (Ω, $\mathcal{F}$, P) does not contain a P-atom, unless P is trivial.

It follows, on pain of triviality, that the range of P, and hence Ω and $\mathcal{F}$, are uncountable. All the more, it is hard to see how rationality requires this of an agent's probability space.

It seems, then, that all four versions of the hypothesis so far considered are untenable. For all that has been said so far, though, a 'tailoring' version restricted to uncountable probability spaces might still survive. Indeed, here we have a positive result due to van Fraassen [55]. Suppose that $\rightarrow$ distributes over $\cap$ and $\cup$, obeys modus ponens and centering, and the principle that $A\rightarrow A = \Omega$. Such an $\rightarrow$ conforms to the logic *CE*. Van Fraassen shows:

CE tenability result: Any probability space can be extended to one for which PCCP holds, with an $\rightarrow$ that conforms to CE.

Of course, the larger space for which PCCP holds is uncountable. He also shows that $\rightarrow$ can have still more logical structure, while supporting PCCP, provided we restrict the admissible iterations of $\rightarrow$ appropriately.

A similar strategy of restriction protects Adams' version of the hypothesis from the negative results. He applies a variant of PCCP to unembedded conditionals of the form A→B, where A and B are conditional-free. More precisely, he proposes:

Adams' Thesis (AT): For an unembedded conditional A→B,

$$P(A \to B) = P(B|A), \text{ if } P(A) > 0,$$
$$= 1, \text{ otherwise.}$$

Since Adams does not allow the assignment of probabilities to Boolean compounds of conditionals, thus violating the closure assumptions of the probability calculus, 'P' is not strictly speaking a probability function (and thus the negative results, which presuppose that it is, do not apply). McGee extends Adams' theory to certain more complicated compounds of conditionals. He later refines (AT) (in [11]), using Popper functions to give a more nuanced treatment of conditionals with antecedents of probability 0. Finally, Stalnaker and Jeffrey (*ibid*) offer an account of the conditional as a random variable. They recover an analogue of (AT), with expectations replacing probabilities, and generalize it to encompass iterations and Boolean compounds of conditionals.

As the recency of much of this literature indicates, this is still a flourishing field of research. The same can be said for virtually all of the points of contact between probability and logic that we have surveyed here.

SUGGESTED FURTHER READING

Space precludes giving bibliographical details for all the authors cited, but where omitted they are easily found elsewhere. Kyburg [29] contains a vast bibliography of the literature on probability and induction pre-1970. Also useful for references before 1967 is the bibliography for "Probability" in the *Encyclopedia of Philosophy*. Earman [10] and Howson and Urbach [22] have more recent bibliographies, and give detailed presentations of the Bayesian program. Skyrms [51] is an excellent introduction to the philosophy of probability. Von Plato [59] is more technically demanding and more historically oriented, with another extensive bibliography that has references to many landmarks in the development of probability theory this century. Fine [13] is still a highly sophisticated survey of and contribution to various foundational issues in probability. Billingsley [2] and Feller [12] are classic textbooks on the mathematical theory of probability. Mückenheim [38] surveys the literature on "extended probabilities" that take values outside the real interval [0, 1]. Eells and Skyrms [11] is a fine collection of articles on probabilities of conditionals. Fenstad (1980) discusses further connections between probability and logic, emphasizing probability functions defined on formal languages, randomness and recursion theory, and non-standard methods.[3]

NOTES

[1] Lebesgue's theory of measure and integration allows a highly sophisticated treatment of various further concepts in probability theory—random variable, expectation, martingale, and so on—all based ultimately on the characterization of the probability of an event as the measure of a set. Important limit theorems, such as the laws of large numbers and the central limit theorem, are beyond the scope of this chapter. The interested reader is directed to references in the bibliographical notes.

[2] Space limitations preclude us from discussing various other important approaches, including Dawid's prequential theory, theories of fiducial probability by authors such as Fisher, Kyburg and Seidenfeld, those based on fuzzy logic, and those based on complexity theory.

[3] This article was written mostly at Cambridge University, and I am grateful to the Philosophy Department and to Wolfson College for the hospitality I was shown there. I also especially thank Jeremy Butterfield, Alex Byrne, Tim Childers, Haim Gaifman, Matthias Hild, Christopher Hitchcock, Colin Howson, Paul Jeffries, Isaac Levi, Vann McGee, Teddy Seidenfeld, Brian Skyrms, Brian Weatherson and Jim Woodward for their very helpful comments.

Bibliography

[1] E. Adams, *The Logic of Conditionals* (D. Reidel, Dordrecht) 1975.

[2] E. Adams, *A Primer of Probability Logic*, (CSLI, Stanford University, Stanford, California), 1998.

[3] P. Billingsley, *Probability and Measure*, 3rd ed., (John Wiley & Sons, New York), 1995.

[4] R. Carnap, *Logical Foundations of Probability*, (University of Chicago Press, Chicago), 1950.

[5] R. Carnap, "Replies and Systematic Expositions" In *The Philosophy of Rudolf Carnap*, P. A. Schilpp, ed., (Open Court, La Salle, Illinois), 1963, 966–998.

[6] [L. J. Cohen, *The Probable and the Provable*, (Clarendon Press, Oxford), 1977.]

[7] B. De Finetti, "Foresight: Its Logical Laws, Its Subjective Sources", translated in [30], 53-118.

[8] B. De Finetti, *Probability, Induction and Statistics*, (John Wiley & Sons, London), 1972.

[9] [P. Diaconis and S. L. Zabell, "Updating Subjective Probability" *Journal of the American Statistical Association*, 77 (1982), 822–30.]

[10] J. Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, (MIT Press, Cambridge, MA), 1992.

[11] E. Eells and B. Skyrms, eds., *Probability and Conditionals*, (Cambridge University Press, Cambridge), 1994.

[12] W. Feller, *An Introduction to Probability Theory and Its Applications,* (John Wiley & Sons, New York), 1968.

J.E. Fenstad, "Logic and Probability" In *Modern Logic — A Survey*, E. Agazzi, ed., (D. Reidel, Dordrecht), 1980, 223-233.

[13] T. Fine, *Theories of Probability*, (Academic Press, New York), 1973.

[14] P. Ghirardato, "Non-additive Measures of Uncertainty: A Survey of Some Recent Developments in Decision Theory" *Rivista Internazionale di Sciencze Economiche e Commerciali* 40 (1993), 253–276.

[15] [R. N. Giere, "Objective Single-Case Probabilities and the Foundations of Statistics" In P. Suppes, *et al*., eds., *Logic, Methodology and Philosophy of Science* IV, (North-Holland, New York), 1973, 467–83.]

[16] M. Goldstein, "The Prevision of a Prevision" *Journal of the American Statistical Assocociation*, 77 (1983), 822–30.

[17] A. Hájek, "Probabilities of Conditionals—Revisited" *Journal of Philosophical Logic* 18 (1989), 423–428.

[18] A. Hájek, "What Conditional Probability Could Not Be", *Synthese*, forthcoming.

[20] [J. Hintikka, “A Two-Dimensional Continuum of Inductive Methods” In *Aspects of Inductive Logic*, J. Hintikka and P. Suppes, eds., (North-Holland, Amsterdam), 1965, 113–32.]

[21] C. Howson, “Theories of Probability” *British Journal of Philosophy of Science*, 46 (1995), 1–32.

C. Howson, “Logic and Probability”, *British Journal of Philosophy of Science*, 48 (1997), 517–531.

[22] C. Howson and P. Urbach, *Scientific Reasoning: The Bayesian Approach*, 2nd ed., (Open Court, Chicago), 1993.

[23] [E. T. Jaynes, “Prior Probabilities” *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics*, SSC-4 (1968), 227–241.]

[24] R. Jeffrey, *The Logic of Decision*, 2nd. ed., (University of Chicago Press, Chicago), 1983.

[25] R. Jeffrey, *Probability and the Art of Judgment*, (Cambridge University Press, Cambridge), 1992.

[26] [D. Kahneman, P. Slovic and A. Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*, (Cambridge University Press, Cambridge), 1982.]

[27] J. G. Kemeny, “Fair Bets and Inductive Probabilities” *Journal of Symbolic Logic*, 20 (1955), 263–273.

[28] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitrechnung*, Ergebnisse Der Mathematik, 1933; transl. as *Foundations of Probability*, (Chelsea Publishing Co., New York), 1950.

[29] H. E. Kyburg, *Probability and Inductive Logic,* (Macmillan, New York), 1970.

[30] H. E. Kyburg, & H. E. Smokler, eds., *Studies in Subjective Probability*, 2nd ed. (Robert E. Krieger Publishing Co., Huntington, New York), 1980.

[31] H. Leblanc, “Alternatives to Standard First-Order Semantics” In *Handbook of Philosophical Logic*, Vol. I., D. Gabbay and F. Guenthner, eds., (D. Reidel, Dordrecht), 1983, 189–274.

[32] [I. Levi, *The Enterprise of Knowledge,* (MIT Press, Cambridge, MA), 1980.]

[33] D. K. Lewis, *Papers in Metaphysics and Epistemology*, (Cambridge University Press, Cambridge), 1999.

[34] D. K. Lewis, "Probabilities of Conditionals and Conditional Probabilities" *Philosophical Review*, 85 (1976), 297–315.

[35] D. K. Lewis, "Probabilities of Conditionals and Conditional Probabilities II" *Philosophical Review*, 95 (1986), 581–589.

[36] D. K. Lewis, "A Subjectivist's Guide to Objective Chance" In *Studies in Inductive Logic and Probability*, Vol II., R. Jeffrey, ed. (University of California Press, Berkeley), 1980, 263–293; reprinted in *Philosophical Papers Volume II,* (Oxford University Press, Oxford).

[37] [P. Mongin, "Some Connections Between Epistemic Logic and the Theory of Nonadditive Probability" In *Patrick Suppes: Scientific Philosopher*, Vol. 1, P. Humphreys, ed. (Kluwer, Dordrecht), 1994, 135–71.]

[38] W. Mückenheim, "A Review of Extended Probabilities" *Physics Reports*, Vol. 133, No. 6 (1986), 337–401.

[39] [J. Paris and A. Vencovská "In Defence of the Maximum Entropy Inference Process" *International Journal of Approximate Reasoning*, 17 (1997), 77–103.]

[40] J. Pearl, "Probabilistic Semantics for Nonmonotonic Reasoning" In *Philosophy and AI: Essays at the Interface*, R. Cummins and J. Pollock, eds., (MIT Press, Cambridge, MA), 1991, 157–188.

[41] [G. Polya, *Patterns of Plausible Inference*, 2nd ed., (Princeton University Press, Princeton, NJ) 1968.]

[42] K. Popper, *The Logic of Scientific Discovery*, (Basic Books, New York), 1959.

[43] [K. Popper, "The Propensity Interpretation of Probability" *British Journal of the Philosophy of Science*, 10 (1959), 25–42.]

[44] F. P. Ramsey, "Truth and Probability" In *Foundations of Mathematics and other Essays*, 1926; reprinted in [30], 23-52.

[45] [H. Reichenbach, *The Theory of Probability,* (University of California Press, Berkeley), 1949.]

[46] A. Renyi, *Foundations of Probability*, (Holden-Day, Inc., Boca Raton), 1970.

[47] L. J. Savage, *The Foundations of Statistics*, (John Wiley & Sons, New York) 1954.

[48] [D. Scott and P. Krauss, "Assigning Probabilities to Logical Formulas" In *Aspects of Inductive Logic*, J. Hintikka and P. Suppes, eds., (North-Holland, Amsterdam), 1966, 219–264.]

[49] G. Shafer, “Constructive Probability” *Synthese*, 48 (1981), 1–60.

[50] B. Skyrms, *Causal Necessity*, (Yale University Press, New Haven), 1980.

[51] B. Skyrms, *Choice and Chance*, 4th ed., (Wadsworth Publishing Company, Belmont, California), 1999.

[52] B. Skyrms, “Dynamic Coherence and Probability Kinematics” *Philosophy of Science*, 54 (1987), 1–20.

[53] [B. Skyrms, “Updating, Supposing, and MAXENT”, *Theory and Decision*, 22 (1987), 225–246.]

[54] R. Stalnaker, “Probability and Conditionals” *Philosophy of Science*, 37 (1970), 64–80.

[55] B. van Fraassen, “Probabilities of Conditionals” In *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, W.L. Harper and C. Hooker, eds., (D. Reidel, Dordrecht), 1976, 261–301.

[56] [B. van Fraassen, “Probabilistic Semantics Objectified I” *Journal of Philosophical Logic*, 10 (1981), 371–394; part II, 495–510.]

[57] [B. van Fraassen, “Gentlemen’s Wagers: Relevant Logic and Probability”, *Philosophical Studies,* 43 (1983), 47–61.]

B. van Fraassen, "Belief and the Problem of Ulysses and the Sirens", *Philosophical Studies*. **77** (1995), 7-37.

[58] R. von Mises, *Probability, Statistics and Truth*, revised English edition, (Macmillan, New York), 1957.

[59] J. von Plato, *Creating Modern Probability*, (Cambridge University Press, Cambridge), 1994.

[60] [P. Walley, *Statistical Reasoning with Imprecise Probabilities*, (Chapman & Hall, London), 1991.]