8

A Philosopher's Guide to Probability

Alan Hájek

Autobiographical Prelude

Once upon a time I was an undergraduate majoring in mathematics and statistics. I attended many lectures on probability theory, and my lecturers taught me many nice theorems involving probability: 'P of this equals P of that', and so on. One day I approached one of them after a lecture and asked him: 'What is this "P" that you keep on writing on the blackboard? *What is probability?*' He looked at me like I needed medication, and he told me to go to the philosophy department.

And so I did. (Admittedly, my route there was long and circuitous.) There I found a number of philosophers asking the very same question: what is probability? All these years later, it's still one of the main questions that I am working on. I still don't feel that I have a completely satisfactory answer, although I like to think that I've made some progress on it. For starters, I know many things that probability is *not*; there are various highly influential analyses of it which cannot be right – we will look at them shortly. As to promising directions regarding what probability *is*, I will offer my best bets at the end.

Introduction

Bishop Butler's dictum that 'probability is the very guide of life' is as true today as it was when he wrote it in 1736 (Butler, 1961). It is almost platitudinous to point out the importance of probability in statistics, physics, biology, chemistry, computer science, medicine, law, meteorology, psychology, economics, and so on. Probability is crucial to any discipline that deals with indeterministic processes, any discipline in which our ability to predict outcomes is imperfect – that is, any serious empirical discipline. Probability is also seemingly ubiquitous outside the academy. Probabilistic judgements of the efficacy and side-effects of a pharmaceutical drug determine whether or not it is approved for release to the public. The fate of a defendant on trial for murder hinges on the jurors' opinions about the probabilistic weight of evidence. Geologists calculate the probability that an earthquake of a certain intensity will hit a given city, and engineers accordingly build skyscrapers with specified probabilities of withstanding such earthquakes. Probability undergirds even measurement itself, since the error bounds that accompany measurements are based on probability. We find probability wherever we find uncertainty – that is, almost everywhere in our lives.

It is surprising, then, that probability is a comparative latecomer on the intellectual scene. To be sure, inchoate ideas about chance date back to antiquity – Epicurus, and later Lucretius, believed that atoms occasionally underwent indeterministic swerves. In the middle ages, Averroes had a notion of 'equipotency' that might be regarded as a precursor to probabilistic notions. But probability theory was not conceived until the 17th century, when the study of gambling games motivated the first serious mathematical study of chance by Pascal and Fermat, culminating in *Port-Royal Logic* (Arnauld and Nicole, 1662). Over the next three centuries, the theory was developed by such authors as Huygens, Bernoulli, Bayes, Laplace, Condorcet, de Moivre, Venn, Johnson and Keynes. Arguably, the crowning achievement was Kolmogorov's (1933) axiomatization, which put probability on a rigorous mathematical footing.

When I asked my professor 'What is probability?', there were two ways to understand that question, and thus there are two kinds of answer that could be given (apart from bemused advice to seek attention from a doctor, or at least a philosopher). First, the question may be understood as: *How should probability theory be formalized?* This is a mathematical question, to which Kolmogorov's axiomatization is the orthodox answer. I review his theory in the next section, and it was given to me at great length in my undergraduate statistics courses. Second, the question may be understood as (and this is more what I intended): *What do statements of probability mean?* This is a philosophical question, and while the mathematical theory of probability certainly bears on it, the answer or answers must come from elsewhere – in my case, from the philosophy department.

The Formal Theory of Probability

More than seventy years old, Kolmogorov's theory of probability is still state-of-the-art for most mathematicians and statisticians. In it, probabilities are numerical values that are assigned to 'events', understood to be certain sets of possibilities belonging to some 'universal set' Ω (the set of all possible outcomes). Probabilities conform to the following axioms. They are non-negative:

$P(X) \geqslant 0.$

Their maximal value is 1, the probability of the universal set:

$P(\Omega) = 1.$

And they are additive: the probability that one of two mutually exclusive events occurs is the sum of their individual probabilities:

$P(X \text{ or } Y) = P(X) + P(Y)$ if X and Y cannot both occur.

For example, for the random experiment of tossing a fair die once and observing how it lands, a natural universal set would be {1, 2, 3, 4, 5, 6}. Each of these outcomes presumably has probability 1/6. The three different ways that the die can land odd (1, 3 and 5) are mutually exclusive, so $P(\text{die lands odd}) = 1/6 + 1/6 + 1/6 = 3/6$.

The *conditional probability of X given Y* (written here as $P(X \mid Y)$) is given by the ratio of unconditional probabilities:

$P(X \mid Y) = P(X \text{ and } Y)/P(Y)$, provided $P(Y) > 0$.

Thus the probability that our fair die lands 1 is 1/6, but the conditional probability that it lands 1, *given* that it lands odd, is 1/3:

$$P(\text{die lands } 1) = 1/6;\ P(\text{die lands } 1 \mid \text{die lands odd}) = \frac{1/6}{3/6} = 1/3$$

A particularly famous result involving conditional probabilities is *Bayes's theorem*. Here's an elegant formulation:

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

More commonly used is:

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

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

where '$\neg A$' symbolizes 'not A'.

If $P(X \mid Y) = P(X)$, then X and Y are said to be *independent*. Intuitively, the occurrence of one of the events is completely uninformative about the occurrence of the other. Thus successive tosses of a coin or successive spins of a roulette wheel are typically regarded as independent. Two cautions: first, the locution 'X is independent of Y' is somewhat careless, encouraging one to forget that independence is a relation that events or sentences bear to a *probability assignment*. Second, this technical sense of 'independence' should not be identified unreflectively with causal independence, or any other pre-theoretical sense of the word, even though such identifications are often made in practice.

Independence plays a central role in probability theory. Many of those theorems that my statistics professors taught me hinge on it – for example the so-called 'laws of large numbers', which formalize what is popularly known as the 'law of averages', the statistical inevitability with which certain processes yield the long run frequencies that one would 'expect'. Think of how a fair coin is very likely to land heads half the time in the long run, assuming that the tosses are independent. If they weren't – if, for example, the coin somehow had a memory of how it behaved previously, and altered its behaviour accordingly – then all bets would be off as far as the laws of large numbers are concerned. Don't laugh – many gamblers act as if chance devices such as coins *do* have memories. For example, after a run of heads, people often feel inclined to bet on tails, thinking that somehow it is 'due'. Assuming that the tosses really are independent, this is an example of the so-called *gambler's fallacy*.

The next section turns to the so-called *interpretations* of probability, attempts to answer the central *philosophical* question: What sorts of things are probabilities? The term 'interpretation' is misleading here. Various quantities that intuitively have nothing to do with 'probability' obey Kolmogorov's axioms – for example length, volume and mass, scaled to assign a value of 1 to some chosen 'maximal' object – and are thus 'interpretations' of it, but not in the intended sense. Nevertheless, we will silence our scruples and follow this common usage in our quick survey. (See Hájek, 2003a, for a more detailed discussion.)

Interpretations of Probability

The *classical* interpretation, historically the first, can be found in the works of Pascal, Huygens, Bernoulli and Leibniz, and it was famously presented by Laplace (1951). Cardano, Galileo and Fermat also anticipated this interpretation. Suppose that our evidence does not discriminate among the members of some set of possibilities – either because that evidence provides equal support for each of them, or because it has no bearing on them at all. Then 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*. This interpretation was inspired by, and typically applied to, games of chance that by their very design create such circumstances – for example the classical probability of a fair die landing with an even number face up is 3/6. Probability puzzles typically take this means of calculating probabilities for granted.

Unless more is said, however, this interpretation yields contradictory results: you have a one-in-a-million chance of winning the lottery; but either you win or you don't, so each of these possibilities has a probability of 1/2! We might look for a 'privileged' partition of the possibilities (into mutually exclusive and exhaustive subsets), but we will not always find one. And even if we do, nothing licenses ruling out biases a priori.

The *logical* interpretation of probability, developed most extensively by Carnap (1950), sees probability as an extension of logic. Traditionally, logic aims to distinguish valid from invalid arguments by virtue of the syntactic form of the premises and conclusion. For example, any argument that has the form:

p
If *p*, then *q*
Therefore, *q*

is valid in virtue of this form. An example of this form would be:

Probability theory is fun.
If probability theory is fun, then you should study it.
Therefore, you should study it.

But the distinction between valid and invalid arguments is not fine enough: many invalid arguments are perfectly good, in the sense that the premises strongly support the conclusion. Carnap described this relation of 'support' or 'confirmation' as the logical probability that an argument's conclusion is true, given that its premises are true. He had faith that logic, more broadly conceived, could also give it a syntactic analysis.

His programme did not succeed. A central problem is that changing the language in which items of evidence and hypotheses are expressed will typically change the confirmation relations between them. Moreover, Goodman (1983) showed that inductive logic must be sensitive to the meanings of words, as syntactically parallel inferences can differ wildly in their inductive strength. For example:

All observed snow is white.
Therefore, all snow is white.

is an inductively strong argument, its premise giving strong support to its conclusion. However:

All observed snow is observed.
Therefore, all snow is observed.

is inductively weak, its premise providing minimal support for its conclusion. It is quite unclear how a notion of logical probability can respect these intuitions.

Frequency interpretations date back to Venn (1876). Gamblers, actuaries and scientists have long understood that relative frequencies bear an intimate relationship to probabilities. Frequency interpretations posit the most intimate relationship of all: identity. Thus, the probability of '6' on a die that lands '6' three times out of ten tosses is, according to the frequentist, 3/10. In general:

> the probability of an outcome *A* in a reference class *B* is the proportion of occurrences of *A* within *B*.

Frequentism is still the dominant interpretation among scientists who seek to capture an objective notion of probability, heedless of anyone's beliefs. It is also the philosophical position that lies in the background of the classical Fisher/Neyman-Pearson approach that is used in most statistics textbooks. Frequentism does, however, face some major objections. For example, a coin that is tossed exactly once yields a relative frequency of heads of either 0 or 1, whatever its bias may be – an instance of the infamous 'problem of the single case'. A coin that is tossed twice can only yield relative frequencies of 0, 1/2, and 1. And in general, a finite number n of tosses can only yield relative frequencies that are multiples of $1/n$. Yet it seems that probabilities can often fall between these values. Quantum mechanics, for example, posits irrational-valued probabilities such as $1/\sqrt{2}$.

Some frequentists (notably Reichenbach, 1949, and von Mises, 1957) address this problem by considering infinite reference classes of hypothetical occurrences. Probabilities are then defined as limiting relative frequencies in suitable infinite sequences of trials. If there are in fact only a finite number of trials of the relevant type, then this requires the actual sequence to be extended to a hypothetical or 'virtual' sequence. This creates new difficulties. For instance, there is apparently no fact of the matter of how the coin in my pocket would have landed if it had been tossed indefinitely – it *could* yield any hypothetical limiting relative frequency that you like.

Moreover, a well-known problem for any version of frequentism is the *reference class problem*: relative frequencies must be relativized to a reference class. Suppose that you are interested in the probability that Collingwood will win its next match. Which reference class should you consult? The class of all matches in Collingwood's history? Presumably not. The class of all recent Collingwood matches? That's also unsatisfactory: it is somewhat arbitrary what counts as 'recent', and some recent matches are more informative than others regarding

Collingwood's prospects. The only match that resembles Collingwood's next match in every respect is that match itself. But then we are saddled again with the problem of the single case, and we have no guidance to its probability in advance.

Propensity interpretations, like frequency interpretations, regard probability as an objective feature of the world. Probability is thought of as a physical propensity, or disposition, or tendency of a system to produce given outcomes. This view, which originated with Popper (1959), was motivated by the desire to make sense of single-case probability attributions on which frequentism apparently foundered, particularly those found in quantum mechanics. Propensity theories fall into two broad categories. According to *single-case* propensity theories, propensities measure a system's tendencies to produce given outcomes; according to *long-run* propensity theories, propensities are tendencies to produce long-run outcome frequencies over repeated trials (see Gillies, 2000, for a useful survey).

Single-case propensity attributions face the charge of being untestable. Long-run propensity attributions may be considered to be verified if the long-run statistics agree sufficiently well with those expected, and falsified otherwise; however, then the view risks collapsing into frequentism, with its attendant problems. A prevalent objection to any propensity interpretation is that it is uninformative to be told that probabilities are 'propensities'. For example, what exactly is the property in virtue of which this coin has a 'propensity' of 1/2 of landing heads (when suitably tossed)? Indeed, some authors regard it as mysterious whether propensities even obey the axioms of probability in the first place.

Subjectivist interpretations – sometimes called 'Bayesian', due to their often being associated with Bayes's theorem – were pioneered by Ramsey (1931) and de Finetti (1937). They see probabilities as *degrees of belief*, or *credences* of appropriate agents. These agents cannot really be actual people, since, as psychologists have repeatedly shown, people typically violate probability theory in various ways, often spectacularly so (although sometimes they may fruitfully be modelled as obeying it). Instead, we imagine the agents to be ideally rational.

But what are credences? De Finetti identifies an agent's subjective probabilities with his or her betting behaviour. For example:

> your probability for the coin landing heads is 1/2 *if and only if* you are prepared to buy or sell for 50 cents a ticket that pays $1 if the coin lands heads, nothing otherwise.

All of your other degrees of belief are analysed similarly.

This analysis has met with many objections. Taken literally, it assumes that opinions would not exist without money, and moreover that you must value money linearly; but if it is just a metaphor, then we are owed an account of the literal truth. Even if we allow other prizes that you value linearly, problems remain: since your behaviour in general, and your betting behaviour in particular, are the result of your beliefs and desires working in tandem, any such proposal fails to resolve these respective components. You may wish to misrepresent your true opinion; or you may particularly enjoy or abhor gambling; or, like a Zen master, you may lack a desire for worldly goods altogether. In each case, your betting behaviour is a highly misleading guide to your true probabilities.

A more sophisticated approach, championed by Ramsey, seeks to fix agents' utilities (numbers that measure how desirable things are according to the agent) and probabilities simultaneously by appeal to their preferences. Suppose that you have a preference ranking of various possible states of affairs and gambles among them, meeting certain conditions required by rationality (for example, if you prefer A to B and B to C, then you prefer A to C). Then we can prove a 'representation' theorem: these preferences can be represented as resulting from an underlying probability distribution and utility function. This approach avoids some of the objections to the betting interpretation, but not all of them. Notably, the essential appeal to gambles again raises the concern that the wrong quantities are being measured. And notice that the representation theorem does not show that rational agents' opinions *must* be represented as probabilities; it merely shows that they *can* be, leaving open that they can also be represented in *other*, substantively different ways.

Radical subjectivists such as de Finetti recognize no constraints on initial, or 'prior', subjective probabilities beyond their conformity to Kolmogorov's axioms. But they typically advocate a learning rule for updating probabilities in the light of new evidence. Suppose that you initially have a probability function $P_{initial}$ and that you become certain of an event E (and of nothing more). What should be your new probability function P_{new}? The favoured updating rule among Bayesians is conditionalization; P_{new} is related to $P_{initial}$ as follows:

(Conditionalization) $P_{new}(X) = P_{initial}(X \mid E)$, provided $P_{initial}(E) > 0$.

Radical subjectivism has been charged with being too permissive. It apparently licenses credences that we would ordinarily regard as crazy. For example, you can assign without its censure a probability of 0.999 to your navel ruling the universe – provided that you remain coherent (and update by conditionalization). Radical subjectivism also seems to allow inferences that are normally considered fallacious, such as the gambler's fallacy (believing, for instance, that after a surprisingly long run of heads, a fair coin is more likely to land tails). Rationality, the objection goes, is not so ecumenical.

A standard defence (see, for example, Savage, 1954; Howson and Urbach, 1993) appeals to famous 'convergence-to-truth' and 'merger-of-opinion' results. Roughly, these say that in the long run, the effect of choosing one prior probability function rather than another is washed out: successive conditionalizations on the evidence will, with probability one, make a given agent eventually converge on the truth, and thus initially discrepant agents eventually come to agreement. Unfortunately, these theorems tell us nothing about how quickly the convergence occurs. In particular, they do not explain the unanimity that we in fact often reach, and often rather rapidly. We will apparently reach the truth 'in the long run'; but, as Keynes quipped, 'in the long run, we shall all be dead'.

Some Recent Developments

In response, certain subjectivists nowadays are more demanding, adding further constraints to their subjectivism. For example, we might evaluate credences according to how closely they match the corresponding relative frequencies: how well 'calibrated' they are. Various subjectivists believe that rational credences are guided by objective chances (perhaps thought of as propensities), so that if a rational agent knows the objective chance of a given outcome, her degree of belief will be the same. There has been important research on the aggregation of opinions and preferences of multiple agents. This problem is well known to aficionados of the risk-assessment literature, which has yet to be mined by philosophers (see Kaplan, 1992).

Recent times have also seen attempts to rehabilitate the classical and logical interpretations, and in particular the principle of indifference. Some 'objective' Bayesians appeal to information theory, arguing that prior probabilities should maximize 'entropy' – a measure of the 'flatness' of a probability distribution – subject to the constraints of a given problem. If there are only a finite number of possible outcomes, the method of maximizing entropy reduces to the principle of indifference (as then the flattest possible distribution simply gives equal probability to each outcome); but the method is more general, handling infinite cases as well.

Probability theory has also been influenced by advances in theories of randomness and complexity theory (see Fine, 1973; Li and Vitanyi, 1997) and approaches to the 'curve-fitting' problem – familiar in the computer science, artificial intelligence and philosophy of science literature – that attempt to measure the simplicity of theories. Influential here have been the 'Akaike information criterion' (see Forster and Sober, 1994), 'minimum description length theory' (see Rissanen, 1998), 'minimum message length theory' (see Wallace and Dowe, 1999), and the 'Bayesian information criterion' (see Kieseppä, 2001).

While Kolmogorov's theory remains the orthodoxy, a host of alternative theories of probability have been developed (see Fine, 1973; Mückenheim et al, 1986). For instance, there has been increased interest in non-additive theories, and there has been lively debate regarding the generalization of additivity to infinite cases. Some authors have proposed theories of primitive conditional probability functions, in which conditional probability replaces unconditional probability as the fundamental concept (see Hájek, 2003b).

And recently a cottage industry has sprung up, responding to the so-called 'Sleeping Beauty problem' (Elga, 2000). This involves a scenario in which someone is put to sleep, and then woken up either once or twice depending on the outcome of a fair coin toss (heads: once; tails: twice). But if she is to be woken up twice, her memory of the first awakening is erased. What probability should she give to heads at the first awakening? 'Halfers' say 1/2, 'thirders' say 1/3, and there are many clever arguments on each side. The paradox has prompted much discussion of so-called *self-locating beliefs* – beliefs about who one is, where one is, or what time it is.

Some Future Avenues of Research

What will future research in the philosophical foundations of probability look like? With appropriate degrees of uncertainty, here are some of my best bets.

I think that there is still much research to be done within a broadly Bayesian framework. I have already mentioned the recent rehabilitation of logical probability, and in particular the principle of indifference (for more, see Bartha and Johns, 2001; Festa, 1993; Maher, 2000 and 2001). As probability spaces are often infinite, this will surely resonate with developments in the theory of infinitesimals, for example within the system of 'surreal numbers' (Conway, 1976). Confirmation theory more generally is being systematically explored and developed by authors such as Fitelson (2007).

Probability theory traditionally presupposes classical set theory/classical logic. There is more work to be done on 'non-classical' probability theory (see Weatherson, 2003, for promising directions). Bayesians may want to enrich their theory of induction to encompass logical/mathematical learning in response to the so-called 'problem of old evidence' (see Zynda, 1995, for a good discussion), and to allow for the formulation of new concepts and theories. Fertile connections between probability and logic have been explored under the rubric of 'probabilistic semantics' or 'probability logic' (see Hailperin, 1996; Adams, 1998). Roeper and Leblanc (1999) develop such probabilistic semantics for primitive conditional probability functions. More generally, I envisage increased attention to the theory of such functions (see, for instance, Festa, 1993, for a treatment of Bayesian confirmation theory that takes such functions as primitive, and Hájek, 2003b, for general arguments in favour of such functions). I expect the highly sophisticated recent investigations of anomalies in the mathematical foundations of conditional probability to continue (see, for example, Seidenfeld et al, 2001).

Further criteria of adequacy for subjective probabilities will be developed – in particular, candidates for playing a role for subjective probability analogous to the role that truth plays for belief. Objective chance seems to be a prime such candidate, and I foresee further study of it. One avenue I find especially promising finds its inspiration in work by Poincaré and has been pursued by a number of researchers at or associated with Stanford University: Suppes, Engel, Keller and Diaconis (see Strevens, 2003). Very roughly, there are certain symmetries in the way macroscopic variables of various systems evolve. These symmetries ensure a statistical robustness that allows one to abstract away from the details of the microscopic variables underlying the systems. Thus one can confidently predict that there will be stable statistics over roulette wheel outcomes, whichever croupiers happen to be spinning them, stable statistics over populations of various ecosystems, whatever the behaviour of the constituent organisms, and so on. In related work, a number of authors are exploring whether objective chance is compatible with underlying determinism (see, for example, Schaffer, 2007).

I expect that non-Bayesian research programmes will also flourish. Non-additive probabilities are getting impetus from considerations of 'ambiguity aversion' (Ghirardato, 2001). Formal learning theory (see Kelly, 2005) is also gaining support, and, more broadly, philosophers will find much interesting work on induction and probabilistic learning in the computer science and artificial intelligence literature. And there is a need for more cross-fertilization between Bayesianism and classical statistics (see, for example, Schervish et al, 2002, for an important recent example of such work). Moreover, in light of work in the economics literature on 'bounded rationality', the study of degrees of incoherence is likely to bear fruit. I foresee related attempts to 'humanize' Bayesianism – for example, the further study of imprecise probability and imprecise decision theory, in which credences need not be precise numbers (see www.sipta.org). And classical statistics, for its part, with its tacit trade-offs between errors and benefits of different kinds, needs to be properly integrated into a more general theory of decision. Meanwhile, the debate among philosophers over the relative merits of 'evidential' and 'causal' decision theory will doubtless continue (see Joyce, 1999).

Probability and decision theory, in turn, will profit from insights in the causal modelling literature. For example, the so-called 'reference class problem' arises because a given event-token can typically be placed under indefinitely many event-types (recall the many different classes in which we could place Collingwood's next match). But progress can be made when the relevant *causes* are identified, and here one can appeal to techniques developed by Spirtes et al (2000), Pearl (2000) and Woodward (2003). These techniques are making quite a splash in biology and the social sciences, and they will be finessed further as a result. More generally, in this brave new world of interdisciplinarity and rapid communication, inferential methods honed within one field are increasingly likely to be embraced by practitioners of another. This is only fitting in the context of this symposium, whose approach to uncertainty has been 'learning from diverse disciplinary and practice approaches'.

Acknowledgements

Many thanks to Ned Hall, who years ago slaved with me on a research proposal that was an ancestor of parts of this chapter, especially for letting me borrow liberally from parts of our chapter (Hájek and Hall, 2002). Many thanks also to Ralph Miles for much-appreciated editorial assistance.

References

Adams, E. (1998) *Probability Logic*, Center for the Study of Language and Information, Stanford University, Palo Alto, CA

Arnauld, A. and Nicole, P. (1662) *La Logique ou L'Art de Penser* [*Logic or the Art of Thinking*, also known as *Port-Royal Logic*], English edition published in 1964, translated by J. Dickoff and P. James, Bobs-Merrill, Indianapolis, IN

Bartha, P. and Johns, R. (2001) 'Probability and symmetry', *Philosophy of Science*, vol 68 (proceedings), ppS109–S122, www.philosophy.ubc.ca/faculty/johns/probsym.pdf, accessed 19 July 2007

Butler, J. (1961, original 1736) *Analogy of Religion*, Frederick Ungar Publishing Company, New York

Carnap, R. (1950) *Logical Foundations of Probability*, University of Chicago Press, Chicago, IL

Conway, J. (1976) *On Numbers and Games*, Academic Press, London

de Finetti, B. (1937) 'La prévision: Ses lois logiques, ses sources subjectives', *Annales de l'Institut Henri Poincaré*, vol 7, pp1–68; translated as de Finetti, B. (1980) 'Foresight: Its logical laws, its subjective sources', in H. E. Kyburg, Jr. and H. E. Smokler (eds) *Studies in Subjective Probability*, Robert E. Krieger Publishing Co., New York

Elga, A. (2000) 'Self-locating belief and the sleeping beauty problem', *Analysis*, vol 60, no 2, pp143–147

Festa, R. (1993) *Optimum Inductive Methods*, Kluwer Academic Publishers, Dordrecht

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

Fitelson, B. (2007) various papers on confirmation theory, http://fitelson.org/research.htm, accessed 19 July 2007

Forster, M. and Sober, E. (1994) 'How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions', *British Journal for the Philosophy of Science*, vol 45, pp1–35

Ghirardato, P. (2001) 'Coping with ignorance: Unforeseen contingencies and non-additive uncertainty', *Economic Theory*, vol 17, pp247–276

Gillies, D. (2000) 'Varieties of propensity', *British Journal for the Philosophy of Science*, vol 51, pp807–835

Goodman, N. (1983) *Fact, Fiction, and Forecast (4th edn)*, Harvard University Press, Cambridge, MA

Hailperin, T. (1996) *Sentential Probability Logic*, Lehigh University Press, Bethlehem, PA

Hájek, A. (2003a) 'Interpretations of probability', in E. N. Zalta (ed) *The Stanford Encyclopedia of Philosophy* (summer 2003 edition), http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/, accessed 19 July 2007

Hájek, A. (2003b) 'What conditional probability could not be', *Synthese*, vol 137, no 3, pp273–323

Hájek, A. and Hall, N. (2002) 'Induction and probability', in P. Machamer and M. Silberstein (eds) *The Blackwell Guide to the Philosophy of Science*, Blackwell, Oxford, UK, pp149–172

Howson, C. and Urbach, P. (1993) *Scientific Reasoning: The Bayesian Approach* (second edition), Open Court, Chicago, IL

Joyce, J. (1999) *The Foundations of Causal Decision Theory*, Cambridge University Press, Cambridge, UK

Kaplan, S. (1992) '"Expert information" versus "expert opinions": Another approach to the problem of eliciting/combining/using expert knowledge in PRA', *Reliability Engineering and System Safety*, vol 35, pp61–72

Kelly, K. (2005) 'Learning, simplicity, truth and misinformation', www.hss.cmu.edu/philosophy/Kelly/papers/Amsterdam.pdf, accessed 19 July 2007

Kieseppä, I. A. (2001) 'Statistical model selection criteria and Bayesianism', *Philosophy of Science*, supplemental volume

Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse der Mathematik*; translated as Kolmogorov, A. N. (1950) *Foundations of Probability*, Chelsea Publishing Company, New York (second English edition 1956)

Laplace, P. S. (1951, original 1814) *A Philosophical Essay on Probabilities*, Dover Publications, New York

Li, M. and Vitanyi, P. (1997) *An Introduction to Kolmogorov Complexity and its Applications* (second edition), Springer-Verlag, New York

Maher, P. (2000) 'Probabilities for two properties', *Erkenntnis*, vol 52, pp63–91

Maher, P. (2001) 'Probabilities for multiple properties: The models of Hesse and Carnap and Kemeny', *Erkenntnis*, vol 55, pp183–216

Mückenheim, W., Ludwig, G., Dewdney, C., Holland, P., Kyprianidis A., Vigier, J., Petroni, N., Bartlett, M. and Jaynes, E. (1986) 'A review of extended probability', *Physics Reports*, vol 133, pp337–401

Pearl, J. (2000) *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge, UK

Popper, K. (1959) *The Logic of Scientific Discovery*, Hutchinson, London

Ramsey, F. P. (1931, original 1926) 'Truth and probability', in R. B. Braithwaite (ed) *Foundations of Mathematics and other Essays*, Routledge and P. Kegan, London, pp156–198; reprinted in H. E. Kyburg, Jr. and H. E. Smokler (eds) (1980) *Studies in Subjective Probability* (second edition), R. E. Krieger, New York, pp23–52; reprinted in D. H. Mellor (ed) (1990) *F. P. Ramsey: Philosophical Papers*, Cambridge University Press, Cambridge, UK

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

Rissanen, J. (1998) 'Hypothesis selection and testing by the MDL principle', *The Computer Journal*, vol 42, no 4, pp260–269

Roeper, P. and Leblanc, H. (1999) *Probability Theory and Probability Logic*, University of Toronto Press, Toronto, Canada

Savage, L. J. (1954) *The Foundations of Statistics*, John Wiley, New York

Schaffer, J. (2007) 'Deterministic chance?', *British Journal for the Philosophy of Science*, vol 58, no 2, pp113–140

Schervish, M. J., Seidenfeld, T. and J. B. Kadane (2002) 'How incoherent is fixed-level testing?', *Philosophy of Science*, vol 69, no 3 (supplement), ppS248–264

Seidenfeld, T., Schervish, M. and Kadane, J. (2001) 'Improper regular conditional distributions', *The Annals of Probability*, vol 29, pp1612–1624

Spirtes, P., Glymour, C. and Scheines, R. (2000) *Causation, Prediction and Search* (second edition), MIT Press, Boston, MA

Strevens, M. (2003) *Bigger Than Chaos*, Harvard University Press, Cambridge, MA

Venn, J. (1876) *The Logic of Chance* (second edition), reprinted 1962, Macmillan, New York

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

Wallace, C. S. and Dowe, D. L. (1999) 'Minimum message length and Kolmogorov complexity', *The Computer Journal* (special issue on Kolmogorov complexity), vol 42, no 4, pp270–283

Weatherson, B. (2003) 'From classical to intuitionistic probability', *Notre Dame Journal of Formal Logic*, vol 44, pp111–123

Woodward, J. F. (2003) *Making Things Happen*, Oxford University Press, Oxford, UK

Zynda, L. (1995) 'Old evidence and new theories', *Philosophical Studies*, vol 77, pp67–95