15

INDUCTIVE LOGIC

The equipossibility account of probability enabled continental workers to operate simultaneously with epistemic and aleatory concepts of probability. But for Leibniz, who invented the idea, it meant far more. It enabled him to conceive of probability theory as an integral part of his metaphysics and epistemology. One by-product of this is that Leibniz anticipated the philosophical programme of J. M. Keynes, Harold Jeffreys and Rudolf Carnap, which has come to be known as inductive logic. We cannot study this in isolation from the rest of Leibniz's philosophy. C. D. Broad has called Leibniz 'the greatest pure intellect whom we have known'. One aspect of his intellect is its attempt to combine every aspect of human knowledge, action, and speculation into an elaborately structured but totally co-ordinated unity. We can start on the outside, with inductive logic, but we shall inevitably be drawn into more central concerns.

Leibniz thought that the science of probability would become a 'new kind of logic' (*P.S.* v, p. 448), but this idea lay dormant until around 1920 when it was revived by Jeffreys and Keynes. Later, in the 1940s, Carnap set to work with a particular approach which has, in recent years, been taken as the sole format for inductive logic. The tenets of this programme can be set out tersely as follows: First, there is such a thing as non-deductive evidence. That is, there may be good reasons for believing *p* which do not logically entail *p*. Second, 'being a good reason for' is a relation between propositions. Third, this relation is to be characterized by a relation between sentences in a suitably formalized language. Fourth, there is an ordering of reasons from good to bad, and indeed a measure of the degree to which *r* is a reason for *p*. Fifth, this measure is autonomous, and independent of anyone's opinion: it is an objective measure of the extent to which *r* is a reason for *p*. Sixth, this measure

is global – it applies to any pair of propositions (*r*, *p*) whatsoever, and not just to some classes of propositions. 'We can always estimate which event under given circumstances can be expected with the highest probability' [*P. S.* vii, p. 188]. Moreover, this global relation is a 'formal' one; it depends solely on the form of the relevant sentences, and not on their content.

It is a matter of some curiosity that no one else was rash enough to emulate these theses until recently, and yet Leibniz believed in them almost as a matter of course. He had grand expectations of his programme. It was part of a scheme for a logical syntax which he called the 'Universal Characteristic'. He prophesied that when it was complete men who disagree would pick up pencils exclaiming 'Let us calculate!' and thereby end contention. Most readers of Leibniz have taken this to be the cry of some alien rationalism which assumes that every issue can be settled by deductive proof. Quite the contrary. Leibniz was not in general speaking of proving propositions but only of finding out which are most probable *ex datis* [*P.S.* vii, p. 200–1].

The 'new kind of logic' has plain enough beginnings. One feature of inductive logic, on which Jeffreys *et al* have insisted so vehemently, is that inductive probabilities are relative to the evidence. We cannot speak of the probability of *p;* we can refer only to the probability of *p* relative to *r*. Leibniz was the only man of his time regularly to declare the relational character of probability judgements. This is a natural consequence of his starting point, namely the law. All legal judgement is *ex datis*, and, as we have seen in Chapter 10, legal reasoning remained Leibniz's paradigm. The new kind of logic was a 'natural jurisprudence'.

A second consequence of this legal point of origin is Leibniz's faith in the objectivity of the probability relation. There may be different assessments of legal evidence, but there is only one *right* assessment. The same goes for probability. Whether or not *r* is a reason for *p* is not a matter of personal opinion. For example, most people thought Copernicus was wrong when he promulgated his heliocentric theory. They thought the available data confuted this unorthodox doctrine. Yet, according to Leibniz, it was still 'incomparably more probable' that Copernicus was right [*N.E.* iv, ii. 14].

With jurisprudence for his model Leibniz thought of probability as relational and objective. This creates the notion of an inductive logic but the next steps are harder. Let us suppose our universe of discourse can be represented by some set of disjoint alternatives (for example, the 2^{100} possible outcomes of tossing a penny 100 times). Let *t* be the empty tautological 'evidence' stating that one of these alternatives will come to pass. If we can assign 'prior probabilities' to the set of alternatives, on the basis of data *t*, then we can work out the 'posterior probability' for any event (say, the event 100th toss gives heads) relative to any data in the universe of discourse (say, 'first 3 tosses gave tails'). In general, an inductive logician can compute the probability of *p*, given *r*, where *p* and *r* are in some universe of discourse *U*, if he knows the prior probability, on tautological 'evidence', of every possible state of affairs expressed in *U*. It has been an aim of Jeffreys and of Carnap to develop such prior distributions for various *U*. Although Jeffreys sticks to realistic *U* and Carnap to simplistic models, they are both moved by two desiderata. One is pragmatic: the resulting posterior probabilities ought, in many cases, to coincide with our unformalized hunches about what is a good reason. The second kind of requirement is 'formal', and consists of conditions of symmetry. For example, suppose that $r(a)$ and $r(b)$ are propositions saying the same thing about individuals *a* and *b* respectively; likewise for $p(a)$ and $p(b)$. Then $r(a)$ should be exactly as good a reason for $p(a)$ as $r(b)$ is for $p(b)$. In the case of prior probabilities relative to tautological 'evidence', a set of possibilities with the same formal structure should be assigned the same prior probability. 'Equal suppositions deserve equal consideration': That, said Leibniz, is the first maxim of probability theory, known even to peasant farmers dividing land [*N.E.* IV, xiv. 9].

Unfortunately it is by no means clear which suppositions are equal. Compare the case of three dice, which first cropped up in Galileo's memorandum. Is the supposition that we get two aces and a deuce 'equal' to the supposition that we get an ace, a deuce and a six? The answer is 'yes' if partitions constitute equal suppositions, but not if permutations within partitions are 'equal'. Now as we noted in Chapter 6, it is a matter of empirical fact that dice, photons and electrons obey, respectively, Maxwell–Boltzmann, Bose–Einstein and Fermi–Dirac statistics, each of which assigns different prior probabilities. In the domain of aleatory probabilities like these it is clear how we are to choose the correct prior: by experiment. But how to choose the right priors for inductive logic? In the simplest case Carnap favoured a probability function c^*

based on the Bose–Einstein statistics for possibilities. It assigns equal probability to each partition. He also set out a continuum of priors, excluding only Maxwell–Boltzmann priors for possibilities because of one pragmatic desideratum – the Maxwell–Boltzmann statistics for possibilities preclude learning from experience. The probability of *p* on any non-entailing evidence *r* is just what it was on the tautology *t*. No prior probability in Carnap's 'continuum of inductive methods' carries any conviction whatsoever. In the case of photons and dice, we can discover prior aleatory probabilities – which are prior propensities – on the basis of experiment. But the programme of inductive logic is supposed to be logical, not only independent of experiment, but the very judge of experiment. How can its prior probabilities be chosen? In modern terms there is no good answer. But Leibniz's metaphysics does provide an answer of sorts.

We have distinguished, on the one hand, the epistemic probability that some possibility is realized, and the aleatory or physical propensity for some possibility to exist. In modern opinion the latter exists, so far as we know, only for a fairly small range of chance set-ups. But in Leibniz's opinion every possibility has a propensity to exist. This is a quite specific and central element of his thought. Many people have heard of the Leibnizian doctrine that this is 'the best of all possible worlds'. Less familiar is his dedicated attempt to explain why there is a world at all, and in particular just this one. According to Leibniz every possible world has some tendency to exist, and our world is the one with the greatest propensity. It is important to compare the familiar aleatory concept of 'ease of making' an outcome with the seemingly different concept of making of a possible world, both summed up in our quotation on page 127 above, by a pun on the word 'feasible'.

As we noted in the last chapter, Leibniz used 'equally possible' to mean something like 'having an equal propensity to occur'. He is also the first to use 'possibility' for a quite different idea which, for today's logician, has superseded every other. 'Possible', according to him, means internally consistent. Leibniz was the first modern to understand that proof is a formal matter, attaching to the form of sentences, not to their content. He defined 'necessarily true' as provable from identities in a finite number of steps. Possibility, then, is freedom from contradiction. For us it follows at once that there are two quite distinct concepts of possibility in question; for consistency is not propensity to occur. But in Leibniz's lists of definitions he regularly explains possibility as freedom from contradiction, and in the next breath speaks of one thing being more possible than another. Something is easy or makeable if it is 'very possible'.

Leibniz's twinned notions of possibility are a deliberate part of his metaphysics. His theory about creation, or what he once called 'the radical origination of things' involves possible objects striving for existence. In the esoteric writings we are invited to contemplate consistent notions having more than mere internal consistency: they have a positive drive to come into being. The more they have of this, the more possible they are. 'The possible demands existence by its very nature, in proportion to its possibility, that is to say, its degree of essence' [*P.S.* VII, p. 194]. In emphasizing this aspect of Leibniz's thought, I admit that I am playing down his other story, which is far better known, in which reality is determined by 'the principle of contingency, or of the existence of things, i.e. of what is or appears the best among several equally possible things' [*P.S.* IV, p. 438].

Leibniz's methodology of science always mirrors his metaphysics, but nowhere is the isomorphism more striking than in the analysis of probability. His contemporaries and predecessors all employ some terminology of the 'facility' of getting an outcome with a die. That, he reminds us, means 'makeable'. All early writers speak of the ease of making various outcomes, but for Leibniz ease of making goes with possibility, and probability is degree of possibility. To repeat the quotations of the preceding chapter, experiments show what is more or less makeable 'in the current state of the world'. What is facile *in re* corresponds to what is probable *in mente*. Note the parallel to the metaphysics. In the esoteric writings we do not read so much of a God choosing among internally consistent state descriptions that which describes a most perfect world: God's role is to conceive the possibilities. The creatability of the things will correspond to the degree of possibility in the divine mind. Similarly, in our world the objective propensities of different outcomes to occur are the foundation of our mental expectations, the probabilities, which, as Leibniz had said, are degrees of possibility. Even for Leibniz such an intertwining of a special science and deepest metaphysics must seem bizarre; so I was glad to find that Dietrich Mahnke [1925] anticipated my interpretation, and took the probability–possibility–facility–creatability nexus as a final proof of

the way that Leibniz linked ontology and physics. Margaret Wilson [1971] has, however, recently contested this point of view.

We now return to our starting point. Leibniz had learned from the law that probability is a relation between hypotheses and evidence. But he learned from the doctrine of chances that probabilities are a matter of physical propensities. Even now no philosopher has satisfactorily combined these two discoveries. Leibniz's combination, although unsatisfactory, is more fascinating than most. On the one hand we have degrees of makeability *in re*, which we may gloss as tendencies to produce stable frequencies. These are the basis of probabilities *in mente*. In particular cases, such a line of thought can be sound. For example if *r* asserts only that in some chance set-up the objective tendency is to produce outcome *E* on repeated trials with stable relative frequency *f*, then the probability of the hypothesis that *E* occurs on the next trial, relative to this data *r*, is surely *f*. Leibniz appears to be inclined to say that this local piece of reasoning has general application. Just as the possible worlds in the mind of God vie with one another for creation, so all the possibilities that we can distinguish will also have some propensity to be actual. We can apply the calculus of probabilities to these possible worlds, using yet another aspect of Leibniz's grand scheme.

Textbooks on probability nowadays often begin with a chapter on combinations and permutations. Inevitably we take Leibniz's youthful *Ars combinatoria* to be in the same line of business. This early monograph on the theory of combinations confirms his claim to have helped advance probability theory. That he had probability theory in mind is proved by internal evidence and also by his proposal to publish Hudde's tables as an appendix. But that is only a small part of the story. The art of combinations was already an established problem area. It was directed not at probability theory but at ideas.

From any vocabulary of ideas we can build other ideas by formal combination of signs. But not any set of ideas will be instructive. One must have the right ideas. Everyone thought that the right ideas would be simple. From an exhaustive set of simple ideas one would generate all possible complex ideas. This is done formally as an operation on signs for the ideas, and this was the point of the art of combinations. Leibniz's immediate predecessors were enraptured with the thought that the world could be understood from a set of signs. Here I have in mind not so much great figures like

Descartes and Spinoza, but a myriad of lesser intellects whose dozens of programmes for universal grammar were motivated by a belief that if only we could uncover in the collective wisdom of mankind a suitable set of ideas, then we would be able to unlock all the secrets of the universe. Universal grammar, which has recently been presented as a key for understanding mind, was in those days a project for understanding all of nature.

Much of Leibniz's intellectual politics is a part of this ferment. His plans for academies and scientific journals intend to co-ordinate knowledge so that we can discover what are the true underlying ideas. Many of his predecessors hoped to uncover an original language preceding Babel. It would encode the true ideas. Leibniz's better plans did not believe in lost innocence but rather in a science and a language that more and more closely correspond to the structure of the universe. His encyclopedia of unified science would collect all present knowledge so we could sift through it for what is fundamental. With the set of ideas that it generated, we could formulate the Universal Characteristic. The art of combinations would enable us to compute all descriptions of possible worlds that could be expressed with that stock of ideas. And the possible worlds so described would all have some propensity to exist.

The Characteristic was supposed to enable us to compute the probabilities of disputed hypotheses relative to the available data; if our Characteristic is founded on simple ideas, then there will be no finitely stateable *a priori* reason that would cause one possible world describable in our language to come to pass rather than another. We thus have a set of alternatives constituting a Fundamental Probability Set to which we can apply a uniform prior probability distribution. The prior distribution is applied not because our set is one among whose alternatives we are ignorant; it is a set such that by metaphysics we know each element has some propensity to exist. Relative to our finite knowledge we may be able to assign only a uniform distribution over possibilities, but we will slowly correct this, and as we learn more our probability assignments will asymptotically tend to a maximum for the real world, i.e. the possibility with the highest actual propensity.

The notion of an asymptotically improving language may sound peculiar to students of Carnap, who writes as if the language, its logic, and its prior probability distribution are fixed. But that is not an essential feature of his theory. Indeed as early as 1932 a fellow

member of the Vienna Circle, Friedrich Waismann, was proposing that language and prior probability should be constantly adjusted as we come to know more about the world. And in Carnap's 1952 *Continuum of Inductive Methods* it is argued that different possible worlds are most efficiently studied by different inductive logics. In the case of Leibniz it is also important to emphasize the role of asymptotic improvement. It is crucial to his metaphysics that no finite analysis is ever complete. In particular no humanly practicable language will enable us to write down an exhaustive classification of possible states of affairs. All possibilities that we can delineate will in fact be complex, and stand for a class of simpler possible worlds. All we can do is estimate propensities for this class, and work hard to make both our classification and our estimates better. We are able to apply probability theory here not because of a principle of indifference applied to an *a priori* language, but rather by a metaphysical ascription of propensities to a classification of possibilities based on learning, scholarship and experiment.

Leibniz's new kind of logic is, then, a compound of three disparate elements. First, there is the doctrine of chances. Secondly, there is a theory of possibility. Some parts of it, original with Leibniz, have become the truisms of our logicians, but the parts most pertinent to probability, though the truisms of yesteryear, are now almost wholly repudiated, and those parts which have their chief role in metaphysics are peculiar to Leibniz. Finally, there is the theory of ideas, a final flourish to the intellectual programme of a preceding era.

Ideas, possibility, and chances create a matrix within which inductive logic could be conceived. They leave open technical questions that have recently perplexed inductive logicians. The problem of choice of initial measure function for prior probabilities is present in what Leibniz proposes, but only after Carnap can we understand it properly. Nevertheless, it is pleasant to note that a Leibnizian ought to like Carnap's preferred c^* which assigns equal probability to what Carnap calls 'structure descriptions'. For Leibniz this could be a methodological consequence of the identity of indiscernibles. Structure descriptions are the finest partitions of possibility that produce descriptions of states of affairs that are distinguishable by the predicates available to a monadic language. A uniform prior distribution over structure descriptions is just c^*.

There is an alternative, however. Leibniz had an important theory of what he called 'architectonic' reasoning. There is deductive *a priori* reasoning, and inductive *a posteriori* reasoning, but there is a middle ground of central importance to science. We favour hypotheses for their simplicity and explanatory power, much as the architect of the world might have done in choosing which possibility to create. The paradigm of this kind of reasoning is our preference for the principle of least time over Descartes' explanation of Snell's (purely inductive) law of refraction [*P.S.* I, p. 195, VII, p. 274]. The Leibnizian might wish his prior probability assignments to conform not to c^* but to Harold Jeffreys' 'simplicity postulate'. Jeffreys needs the simplicity postulate in order to get positive probabilities for law-like propositions, but his is a purely epistemological thesis for which only pragmatic reasons have been given. For Leibniz, in contrast, it is one more pleasant consequence of metaphysics. Simplicity of covering laws and variety of phenomena are the twin measures of perfection for possible worlds. Hence laws with those features will have a greater objective tendency to reality than cumbersome or restricted principles. As in our earlier discussion, high probabilities derived from a simplicity postulate are grounded on a metaphysical ascription of propensities.

We should go no further in reconstructing a Leibnizian theory of probability and inductive logic. Its most notable feature is that there is an objectively correct prior distribution of probability for a set of possibilities. The correct distribution is the one that corresponds to the propensity of the possibility to exist – very much as in dice rolling. I doubt that anyone will accept such a Leibnizian foundation for inductive logic. Still, I prefer it to more recent theories of global inductive logic, which have no foundation at all.

16

THE ART OF CONJECTURING

Jacques Bernoulli's *Ars conjectandi* presents the most decisive conceptual innovations in the early history of probability. The author died in 1705. He had been writing the book off and on for twenty years. Although the chief theorem was proved in 1692, he was never satisfied and he never published. The work was finally given to the printer by his nephew Nicholas, and appeared in Basle in 1713. In that year probability came before the public with a brilliant portent of all the things we know about it now: its mathematical profundity, its unbounded practical applications, its squirming duality and its constant invitation for philosophizing. Probability had fully emerged.

The chief mathematical contribution of the book is plain enough: the first limit theorem of probability. This result has rightly been given the seal of A. N. Kolmogorov as being proven by Bernoulli with 'full analytical rigour' [Maistrov, 1974, p. 75]. But what the theorem means is another question. We shall try to find out in the next chapter. First it is worth investigating Bernoulli's own conception of probability. Since we still lack universal agreement on the analysis of probability no one writes dispassionately about the man. He has been fathered with the first subjective conception of probability. Yet Richard von Mises [1951] could cast him as a stalwart frequentist. More recent statisticians such as A. P. Dempster [1966] say he anticipated Jerzy Neyman's approach to inference via confidence intervals. P. M. Boudot [1967] has argued that Bernoulli was a good inductivist and anticipated the theories of Rudolf Carnap. Since each of these approaches to probability is customarily deemed inconsistent with every other, each school claims Bernoulli as its own. The truth of the matter must be that he was, like so many of us, attracted to all these seemingly incompatible ideas and was unsure where to rest his case.