S. L. ZABELL

SYMMETRY AND ITS DISCONTENTS

The following paper consists of two parts. In the first it is argued that Bruno de Finetti's theory of subjective probability provides a partial resolution of Hume's problem of induction, if that problem is cast in a certain way. De Finetti's solution depends in a crucial way, however, on a symmetry assumption – exchangeability – and in the second half of the paper the broader question of the use of symmetry arguments in probability is analyzed. The problems and difficulties that can arise are explicated through historical examples which illustrate how symmetry arguments have played a important role in probability theory throughout its development. In a concluding section the proper role of such arguments is discussed.

1. THE DE FINETTI REPRESENTATION THEOREM

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of 0,1-valued random variables, which may be thought of as recording when an event occurs in a sequence of repeated trials (e.g., tossing a coin, with 1 if heads, 0 if tails). The sequence is said to be *exchangeable* if all finite sequences of the same length with the same number of ones have the same probability, i.e., if for all positive integers n and permutations σ of $\{1, 2, 3, \ldots, n\}$,

$$P[X_1 = e_1, X_2 = e_2, \ldots, X_n = e_n] = P[X_1 = e_{\sigma(1)}, X_2 = e_{\sigma(2)}, \ldots, X_n = e_{\sigma(n)}],$$

where e_i denotes either a 0 or a 1. For example, when $n = 3$, this means that

$$P[1, 0, 0] = P[0, 1, 0] = P[0, 0, 1] \quad \text{and}$$
$$P[1, 1, 0] = P[1, 0, 1] = P[0, 1, 1].$$

(Note, however, that $P[1, 0, 0]$ is not assumed to equal $P[1, 1, 0]$; in general, these probabilities may be quite different.)

In 1931 the Italian probabilist Bruno de Finetti proved his famous

Brian Skyrms and William L. Harper (eds.), Causation, Chance, and Credence, Vol. I, 155—190.

DE FINETTI REPRESENTATION THEOREM. *Let $X_1, X_2, X_3, \ldots$ be an infinite exchangeable sequence of 0,1-valued random variables, and let $S_n = X_1 + X_2 + \cdots + X_n$ denote the number of ones in a sequence of length n. Then it follows that:*

1. *the limiting frequency $Z =: \lim_{n\to\infty}(S_n/n)$ exists with probability 1.*
2. *if $\mu(A) =: P[Z \in A]$ is the probability distribution of Z, then*

$$P[S_n = k] = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, d\mu(p)$$

for all n and k.[1]

This remarkable result has several important implications. First, contrary to popular belief, subjectivists clearly believe in the existence of infinite limiting relative frequencies – at least to the extent that they are willing to talk about an (admittedly hypothetical) infinite sequence of trials.[2] The existence of such limiting frequencies follows as a purely mathematical consequence of the assumption of exchangeability.[3] When an extreme subjectivist such as de Finetti denies the existence of objective chance or physical probability, what is really being disputed is whether limiting frequencies are objective or physical properties.

There are several grounds for such a position, but all center around the question of what "object" an objective probability is a property of. Surely not the infinite sequence, for that is merely a convenient fiction (Jeffrey 1977). Currently the most fashionable stance seems to be that objective probabilities are a *dispositional property* or *propensity* which manifests itself in, and may be measured with ever-increasing accuracy by, finite sequences of ever-increasing length (see, e.g., Kyburg 1974).

But again, a property of what? Not the coin, inasmuch as some people can toss a so-called "fair" coin so that it lands heads 60% of the time or even more (provided the coin lands on a soft surface such as sand rather than a hard surface where it can bounce). Some philosophers attempt to evade this type of difficulty by ascribing propensities to a *chance set-up* (e.g., Hacking 1965): in the case of coin-tossing, the coin *and* the manner in which it is tossed. But if the coin were indeed tossed in an identical manner on every trial, it would always come up heads or always come up tails; it is precisely because the manner in which the coin is tossed on each trial is *not* identical that the coin can come up both ways. The suggested chance set-up is in fact nothing other than a sequence of objectively differing trials which we are subjectively unable to distinguish between. At best, the infinite limiting frequency is a property of an "object" enjoying both objective and subjective features.

2. DE FINETTI VANQUISHES HUME

The most important philosophical consequence of the de Finetti representation theorem is that it leads to a solution to *Hume's problem of induction:* why should one expect the future to resemble the past? In the coin-tossing situation, this reduces to: in a long sequence of tosses, if a coin comes up heads with a certain frequency, why are we justified in believing that in future tosses of the same coin, it will again come up heads (approximately) the same fraction of the time?

De Finetti's answer to this question is remarkably simple. Given the information that in n tosses a coin came up heads k times, such data is incorporated into one's probability function via

Bayes's rule of conditioning: $P[A \mid B] = P[A \text{ and } B]/P[B]$.

If n is large and $p^* = k/n$, then – except for certain highly opinionated, eccentric, or downright kinky "priors" $d\mu$ – it is easy to prove that the resulting posterior probability distribution on p will be highly peaked about p^*; that is, the resulting probability distribution for the sequence of coin tosses looks approximately like (in a sense that can be made mathematically precise) a sequence of independent and identically distributed Bernoulli trials with parameter p^* (i.e., independent tosses of a p^* coin). By the weak law of large numbers it follows that, with high probability, subsequent tosses of the coin will result in a relative frequency of heads very close to p^*.

Let us critically examine this argument. Mathematically it is, of course, unassailable. It implicitly contains, however, several key suppositions:

1. P is operationally defined in terms of betting odds.
2. P satisfies the axioms of mathematical probability.
3. P is modified upon the receipt of new information by Bayesian conditioning.
4. P is assumed to be exchangeable.

In de Finetti's system, degree of belief is quantified by the betting odds one assigns to an event. By a Dutch book or coherence argument, one deduces that these betting odds should be consistent with the axioms of mathematical probability. Conditional probabilities are initially defined in terms of conditional bets and Bayes's rule of conditioning is deduced as a consequence of coherence. The relevance of conditional probabilities to inductive inference is the *dynamic assumption of Bayesianism* (Hacking 1967): if one learns that B has occurred, then one's new probability assignment is $P[A \mid B]$. In general, however, conditional probabilities can behave in very nonHumeian ways, and (infinite) exchangeability is taken as describing the special class of situations in which Humeian induction is appropriate.

This paper will largely concern itself with the validity of this last assumption. Suffice it to say that, like Ramsey (1926), one may view the subjectivist interpretation as simply capturing *one* of the many possible meanings or useages of probability; that the Dutch book and other derivations of the axioms may be regarded as plausibility arguments (rather than normatively compelling); and that although a substantial literature has emerged in recent decades concerning the limitations of Bayesian conditioning, the difficulties discussed and limitations raised in that literature do not seem particularly applicable to most of the situations typically envisaged in discussions of Hume's problem.

The assumption of exchangeability, however, seems more immediately vulnerable. Isn't it essentially circular, in effect assuming what one wishes to prove? Of course, in one sense this must obviously be the case. All mathematics is essentially tautologous, and any implication is contained in its premises. Nevertheless, mathematics has its uses. Formal logic and subjective probability are both theories of consistency, enabling us to translate certain assumptions into others more readily palatable.

What de Finetti's argument really comes down to is this: if future outcomes are viewed as exchangeable, i.e., no one pattern is viewed as any more or less likely than any other (with the same number of successes), then when an event occurs with a certain frequency in an initial segment of the future, we must, if we are to be consistent, think it likely that that event will occur with approximately the same frequency in later trials. Conversely, if we do not accept this, it means that we must have – prospectively – thought certain patterns more likely than others. Which means that we must have possessed more information than is ordinarily posited in discussions of Humeian induction.

And there the matter would appear to stand. Or does it?

3. THE INSIDIOUS ASSUMPTION OF SYMMETRY

Exchangeability is one of many instances of the use of symmetry arguments to be found throughout the historical development of mathematical probability and inductive logic. But while such arguments often have a seductive attraction, they also often carry with them "hidden baggage": implications or consequences, sometimes far from obvious, which later cast serious doubt on their validity. We will discuss three historically important examples, all involving attempts to justify induction by the use of probability theory, and all (in effect) involving the appropriate choice of prior $d\mu$ in the de Finetti representation.

EXAMPLE 3.1: *Bayes's argument for the Bayes–Laplace prior.*

Consider "an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it" (Bayes 1764). Implicitly invoking a symmetry argument, Bayes argued that "concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another," i.e., that in a sequence of n trials one's probability assignment for S_n, the number of heads, should satisfy

Bayes's Postulate: $P[S_n = k] = 1/(n + 1)$.

That is, the number of heads can assume any of the $n + 1$ values 0, 1, 2,..., n and, absent further information, all $n + 1$ values are viewed as equally likely. In a famous Scholium, Bayes concluded that if this were indeed the case, then the prior probability $d\mu(p)$ must be the "flat" prior dp.[4]

Although Bayes's exact reasoning at this point is somewhat unclear, it can easily be made rigorous: Taking $k = 0$ in the de Finetti representation and using Bayes's postulate, it follows that

$$\int_0^1 p^n \, d\mu(p) = 1/(n + 1).$$

The integral on the left-hand side is the n-th moment of $d\mu$, so Bayes's assumption uniquely determines the moments of $d\mu$. But since $d\mu$ is concentrated on a compact set, it follows by a theorem of Hausdorff that $d\mu$, if it exists, is in turn determined by its moments. That is, there can be at most one probability measure $d\mu$ which satisfies Bayes's assumption $P[S_n = k] = 1/(n + 1)$. But the flat measure dp does satisfy this integral equation, i.e.,

$$\int_0^1 p^n \, dp = 1/(n + 1),$$

hence $d\mu$ must be dp.

Bayes's argument is quite attractive. A modern-day subjectivist might view Bayes's assumption as a *definition* (possibly one of many) of "complete ignorance" (rather than consider "complete ignorance" to be an *a priori* meaningful concept), but would probably find Bayes's argument otherwise unobjectionable.

The argument in its original form, however, did not go uncriticized. As Boole (1854, pp. 369–375) noted, rather than consider the events $[S_n = k]$ to be equally likely, one could equally plausibly take all sequences of a fixed length (or "constitutions") to be so. Thus, for $n = 3$

$$\begin{aligned} P[000] &= P[100] = P[010] = P[001] = P[110] \\ &= P[101] = P[011] = P[111] = 1/8. \end{aligned}$$

To many, this assignment seemed a far more natural way of quantifying ignorance than Bayes's.

Unfortunately, it contains a time-bomb with a very short fuse. As Carnap (1950, p. 565) later noted (and Boole himself had already remarked), this probability assignment corresponds to *independent* trials, and thus remains unchanged when conditioned on the past, an obviously unsatisfactory choice for modeling inductive inference, inasmuch as "past experience does not in this case affect future expectation" (Boole 1854, p. 372).

In his *Logical Foundations of Probability* (1950), Carnap announced that in a later volume, "a quantitative system of inductive logic" would be constructed, based upon a function Carnap denoted c^*. Carnap's c^* function was, in effect, the one already proposed by Bayes. But

Carnap grew uneasy with this unique choice, and in his monograph *The Continuum of Inductive Methods* (1952), he advocated instead the use of a one-parameter family containing c^*. Unknown to Carnap, however, he had been anticipated in this, almost a quarter of a century earlier, by the English philosopher William Ernest Johnson.

EXAMPLE 3.2: *W. E. Johnson's sufficientness postulate.*

In 1924 Johnson, a Cambridge logician, proposed a multinomial generalization of Bayes's postulate. Suppose there are $t \geqslant 2$ categories or types, and in n trials there are n_1 outcomes of the first type, n_2 outcomes of the second type, . . . , and n_t outcomes of the t-th type, so that $n = n_1 + n_2 + \cdots + n_t$. The sequence $(n_1, n_2, \ldots, n_t)$ is termed an *ordered t-partition of n*. Bayes had considered the case $t = 2$, and his postulate is equivalent to assuming that all ordered 2-partitions $(k, n - k)$ are equally likely. Now Johnson proposed as its generalization

Johnson's combination postulate: Every ordered t-partition of n is equally likely.

For example, if $t = 3$ and $n = 4$, then there are 15 possible ordered 3-partitions of 4, viz.:

n_1	n_2	n_3
4	0	0
3	1	0
3	0	1
2	2	0
2	1	1
2	0	2
1	3	0
1	2	1
1	1	2
1	0	3
0	4	0
0	3	1
0	2	2
0	1	3
0	0	4

and each of these is assumed to be equally likely.

Johnson did not work with integral representations but, like Carnap, with finite sequences. In so doing he introduced a second postulate, his "permutation postulate." This was none other the assumption of exchangeability, thus anticipating de Finetti (1931) by almost a decade! (If one labels the types or categories with the letters of a t-letter alphabet, exchangeability here means that all words of the same length, containing the same number of letters of each type, are equally likely). Together, the combination and permutation postulates uniquely determine the probability of any specific finite sequence. For example, if one considers the fifth partition in the table above, $4 = 2 + 1 + 1$, then there are twelve sequences which give rise to such a partition, viz.

x_1	x_2	x_3	x_4
1	1	2	3
1	1	3	2
1	2	1	3
1	2	3	1
1	3	1	2
1	3	2	1
2	1	1	3
2	1	3	1
2	3	1	1
3	1	1	2
3	1	2	1
3	2	1	1

and each of these are thus assumed to have probability $(1/15)(1/12) = 1/180$. The resulting probability assignment on finite sequences is identical with Carnap's c^*.

Despite its mathematical elegance, Johnson's "combination postulate" is obviously arbitrary, and Johnson was later led to substitute for it another, more plausible one, his "sufficientness postulate." This new postulate assumes for all n

Johnson's sufficientness postulate:

$$P[X_{n+1} = j \mid X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n] = f(n_j, n).$$

That is, the conditional probability that the next outcome is of type j depends only on the number of previous trials and the number of previous outcomes of type j, but not on the frequencies of the other types or the specific trials on which they occurred. If, for example $t = 3$, $n = 10$, and $n_1 = 4$, the postulate asserts that on trial 11 the (conditional) probability of obtaining a 1 is the same for all sequences containing four 1's and 6 not -1's, and that this conditional probability does not depend on whether there were six 2's and no 3's, or five 2's

and one 3, and so on. (Note that the postulate implicitly assumes that all finite sequences have positive probability, so that the conditional probabilities are well-defined.)

Johnson's sufficientness postulate makes what seems a minimal assumption: absence of knowledge about different types is interpreted to mean that information about the frequency of one type conveys no information about the likelihood of other types occurring. It is therefore rather surprising that it follows from the postulate that the probability function P is uniquely determined up to a constant:

THEOREM. (Johnson 1932): *If P satisfies the sufficientness postulate and $t \geqslant 3$, then there exists a $k > 0$ such that*

$$f(n_i, n) = \{n_i + k\}/\{n + tk\}.$$

This is, of course, nothing other than Carnap's "continuum of inductive methods."[5]

The de Finetti representation theorem can be generalized to a much wider class of infinite sequences of random variables than those taking on just two values (see, e.g., Hewitt and Savage 1955). In the multinomial case now being discussed, the de Finetti representation states that every exchangeable probability can be written as a mixture of multinomial probabilities. Just as Bayes's postulate implied that the prior $d\mu$ in the de Finetti representation was the flat prior, Johnson's theorem implies that the mixing measure $d\mu$ in the de Finetti representation is the *symmetric Dirichlet prior*

$$p_1^{k-1} p_2^{k-1} \ldots p_t^{k-1} \, dp_1 \, dp_2 \ldots dp_{t-1}:$$

a truly remarkable result, providing a subjectivistic justification for the use of the mathematically attractive Dirichlet prior.[6]

Despite its surface plausibility, Johnson's sufficientness postulate is often too strong an assumption. While engaged in cryptanalytic work for the British government at Bletchley Park during World War II, the English logician Alan Turing realized that even if one lacks specific knowledge about individual category types, the frequencies $n_1, n_2, \ldots, n_t$ may contain relevant information about predictive probabilities, namely the information contained in the *frequencies of the frequencies*.

Let a_r = the number of frequencies n_i equal to r; a_r is called the frequency of the frequency r. For example, if $t = 4$, $n = 10$, and one observes the sequence 4241121442, then $n_1 = 3$, $n_2 = 3$, $n_3 = 0$, $n_4 = 4$

and $a_0 = 1$, $a_1 = 0$, $a_2 = 0$, $a_3 = 2$, $a_4 = 1$. (A convenient shorthand for this is $0^1 1^0 2^0 3^2 4^1$.) Although it is far from obvious, the a_r may be used to estimate cell probabilities: see Good (1965, p. 68).[7]

EXAMPLE 3.3: *Exchangeability and Partial Exchangeability.*

Given the failure of such attempts, de Finetti's program must be seen as a further retreat from the program of attempting to provide a unique, quantitative account of induction. Just as Johnson's sufficientness postulate broadened the class of inductive probabilities from that generated by the Bayes–Laplace prior to the continuum generated by the symmetric Dirichlet priors, so de Finetti extended the class of possible inductive probabilities even further to include *any* exchangeable probability assignment.

But what of the symmetry assumption of exchangeability? Even this is not immune to criticism (as de Finetti himself recognized). Consider the following sequence: 000101001010100010101001.... Scrutiny of the sequence reveals the interesting feature that although every 0 is followed by a 0 or 1, every 1 is invariably followed by a 0. If this feature were observed to persist over a long segment of the sequence (or simply that 1's were followed by 0's with high frequency), then this would seem relevant information that should be taken into account when calculating conditional, predictive probabilities. Unfortunately, exchangeable probabilities are useless for such purposes: if P is exchangeable, then the conditional probabilities

$$P[X_{n+1} = j \mid X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n]$$

depend solely on the number of 1's, and not on their order within the sequence. Thus, exchangeability, despite its plausibility, rules out a natural form of inductive inference and can only be considered valid when "order effects" are ruled out (as, for example, in coin-tossing).

An appropriate generalization of exchangeability that takes such order information into account is the concept of *Markov exchangeability:* all sequences with the same initial letter and the same transition counts (t_{ij} =: number of transitions from state i to state j) are assumed equally likely. Here too a de Finetti representation is possible (Diaconis and Freedman 1980b, 1980c): now one mixes on the possible transition matrices p_{ij}.

Once one has come this far, of course, it is easy to recognize that order effects of this type are merely one of many possible patterns that may be judged to provide useful information, each pattern requiring a corresponding generalization of exchangeability to incorporate the information it provides. To deal with such situations, de Finetti introduced in 1938 the notion of *partial exchangeability* (Diaconis and Freedman 1980c). Although partial exchangeability is an active field of current mathematical research still undergoing development (see, e.g., Diaconis and Freedman 1985), the general outline of the theory is clear: to each pattern corresponds a statistic or symmetry, a representation theorem, and a corresponding mode of inductive inference.

Thus, de Finetti's resolution of Hume's problem of induction is a highly qualified one: it is a theory of coherence. Every person's probability function will contain some symmetry involving past and future, and coherence dictates that patterns observed in the past will be expected to recur in the future.

Despite its highly qualified nature, the above analysis has an important payoff: it demonstrates that Hume's problem is in fact ill-posed; to ask "*why* should the future be expected to resemble the past?" presupposes having already answered the question "*how* is the future expected to resemble the past?" (It is essentially this point that is behind Nelson Goodman's "grue" paradox.) It is a strength of the subjectivist analysis that this point emerges as natural and obvious; indeed, it is essentially forced on one; and to the extent that one can state precisely the ways in which the past and future are conjectured to correspond, it gives a satisfactory solution to Hume's problem.

The successive attempts of Bayes, Johnson, and de Finetti to solve the problem of induction are marked by the invocation of progressively weaker symmetry assumptions. Symmetry, however, has played not only a key role in the the attempts to quantify induction, it has played a central role in the birth and evolution of probability theory, more central perhaps than sometime recognized. In the next three sections it will be argued that the birth of mathematical probability marked a key change in the way symmetry arguments were used; that the early dependence on symmetry arguments to quantify probability, while crucial to its mathematical development, blurred important epistemological distinctions; and that it was only with the challenging of precisely those symmetry arguments in the 19th century that the conceptual clarification of probability became possible.

4. OU MALLON

The simplest and oldest of such arguments is the use of physical or epistemic symmetry to identify a *fundamental probability set* or FPS, i.e., a partition of the space of possible outcomes into equiprobable alternatives. The recognition and use of such sets to compute numerical probabilities for complex events was a key step in the birth of mathematical probability. Once the ability to calculate probabilities in this simple case had been mastered, the outlines of the mathematical theory discerned, and its practical utility recognized, all else followed. Why were the mathematicians of the 17th century able to take this step, while the Greeks, despite their mathematical prowess and penchant for games of chance, were not? The crucial point to recognize is that while for the pioneers of the modern theory the equipossible elements of an FPS were *equally likely*, for the Greeks *none were possible*.

This was because of what G. E. L. Owen has described as "a very Greek form of argument" (Owen 1966), a form of reasoning employed by the Greeks that Leibniz was very fond of and which he called the *principle of sufficient reason*: "for every contingent fact there is a reason why the fact is so and not otherwise ..." (Broad 1975, p. 11). In the words of Leucippus (the only complete sentence of his which has come down to us), "Nothing occurs at random, but everything for a reason and by necessity" (Kirk and Raven 1957, p. 413). Two famous examples will illustrate its use:

4.1. *Anaximander and the position of the earth.* Anaximander (c. 610–540 B.C.), one of the early pre-Socratic Greek philosophers, believed the Earth lay at the center of the universe. But unlike Thales before him, who thought the Earth floated on water, and Anaximenes after, who thought it floated on air, Anaximander thought the Earth was unsupported and remained at the center for reasons of symmetry (*omoiotes*; variously translated as "similarity," "indifference," "equilibrium," or "equiformity").[8] Unfortunately, the text of Anaximander has not survived, and we are dependent on secondary, incomplete, and contradictory later accounts for information about the precise nature of his astronomical beliefs.[9] Our best source is perhaps Aristotle, who reports:

> There are some who say, like Anaximander among the ancients, that [the earth] stays still because of its equilibrium. For it behoves that which is established at the center, and is equally related to the extremes, not to be borne one whit more either up or down or to the sides; and it is impossible for it to move simultaneously in opposite directions, so that it stays fixed by necessity. [*de Caelo* 295 b10]

How closely this reproduces Anaximander's own logic, the exact meaning to be attached to *omoiotes*, indeed the precise nature of the argument itself, is unclear. Nevertheless, the gist of the argument is clearly an appeal to symmetry: for every direction there is an opposite; since there is no more reason for the earth to move in one direction than another, the proper conclusion is that it moves in *neither*.

Although Aristotle expressed scepticism about such reasoning, it was fully accepted by Plato:

> I am therefore persuaded that, in the first place, since the earth is round and in the middle of the heaven, it has not need either of air of any other necessity in order not to fall, but the similarity of the heaven to itself in every way and the equilibrium of the earth suffice to hold it still. For an equilibrated thing set in the midst of something of the same kind will have no reason to incline in one direction more than in another. But as its relationship is symmetrical it will remain unswervingly at rest. [*Phaedo* 108e–109a; c.f. *Timaeus* 62d.12]

4.2. *Parmenides and the creation of the universe.* Parmenides gave a similar argument to show that the universe had never been created:

> And what need would have driven it on to grow, starting from nothing, at a later time rather than an earlier? [Kirk and Raven 1957, p. 273]

Again this is essentially a symmetry argument: if the universe had been created, it must have been at some specific time; inasmuch as there is no more reason for it to have been created at any one time than any other, all possible times are thereby ruled out. Obviously the argument makes some presuppositions, but it had great appeal to Leibniz and appears in his correspondence with Clarke.[10]

It is, as G. E. L. Owen notes,

> a very Greek pattern of argument.... Aristotle retailored the argument to rebut the probability of motion in a vacuum; the Academy adapted it to show that, since no physical sample of equality has more right to serve as a standard sample than any other, the standard sample cannot be physical. And Leibniz found an excellent example in Archimedes's mechanics.... [Owen 1966]

The Greek Pyrrhonian skeptics made systematic use of a similar device for destroying belief. Their goal was to achieve a state of *epoche*, or suspension of judgement about statements concerning the external world, which they believed would in turn lead to *ataraxia*, a state of tranquility, "... saying concerning each individual thing that it no more [*ou mallon*] is than is not, or that it both is and is not, or that it neither is nor is not."[11]

How can *epoche* be achieved? According to Sextus Empiricus (*Outlines of Pyrrhonism* 1.8):

Scepticism is an ability which sets up antitheses among appearances and judgments in any way whatever: by scepticism, on account of the 'equal weight' which characterizes opposing states of affairs and arguments, we arrive first at 'suspension of judgment', and second at 'freedom from disturbance'.

For example, knowledge of what is good is impossible, for what one person thinks good, another may think bad, and

if we say that not all that anyone thinks good is good, we shall have to judge the different opinions; and this is impossible because of the equal validity of opposing arguments. Therefore the good by nature is impossible.

It is important to understand the implications of asserting "*ou mallon*." One might interpret it in a positive sense: although certain knowledge is ruled out, the information we possess is equally distributed between two or more possibilities, and hence we have an equal degree of belief in each. That this was *not* the skeptical position is clear from a passage in Diogenes Laertius (*Life of Pyrrho* 9.74–76):

Thus by the expression "We determine nothing" is indicated their state of even balance; which is similarly indicated by the other expressions, "Not more (one thing than another)," "Every saying has its corresponding opposite," and the like. But "Not more (one thing than another)" can also be taken positively, indicating that two things are alike; for example, "The pirate is no more wicked than the liar." But the Sceptics meant it not positively but negatively, as when, in refuting an argument, one says, "Neither had more existence, Scylla or the Chimaera ..." Thus, as Timon says in the *Pytho*, the statement [*ou mallon*] means just absence of all determination and withholding of assent. The other statement, "Every saying, etc.," equally compels suspension of judgment; *when facts disagree, but the contradictory statements have exactly the same weight, ignorance of the truth is the necessary consequence.* [Emphasis added]

Pyrrhonian skepticism is an extreme position, and the later Academic skeptics developed a theory that combined skepticism about certain knowledge with a description of rational decision based on

probable knowledge.[12] Under Carneades this theory included a scale of the varying degrees of conviction conveyed by an impression, depending on whether it was "credible," "credible and consistent," or "credible, consistent, and tested." Carneades's theory amounts to an early account of qualitative or comparative subjective probability, and one might expect that a later skeptic would go the final step and attempt to numerically measure or describe such degrees of conviction. That this did not happen, it may be argued, was a consequence of the *ou mallon* viewpoint. Witness Cicero's statement:

If a question be put to [the wise man] about duty or about a number of other matters in which practice has made him an expert, he would not reply in the same way as he would if questioned as to whether the number of the stars is even or odd, and say that he did not know; for in things uncertain there is nothing probable [*in incertis enim nihil est probabile*], but in things where there is probability the wise man will not be at a loss either what to do or what to answer. [Cicero *Academica* 2.110]

A 19th century enthusiast of the principle of insufficient reason would have little hesitation in assigning equal probabilities to the parity of the number of stars; this passage thus strikingly illustrates a chasm that had to be crossed before numerical probabilities could be assigned. Cicero was familiar with a theory of probability, indeed much of the *Academica* is devoted to a discussion of Academic probabilism and is one of our major sources of information about it. But for Cicero the probable was limited in its scope, limited in a way that precluded its quantification. The FPS was the basic setting for the early development of mathematical probability – but for Cicero it was a setting in which the very notion of probability itself was inapplicable.

Support for this thesis may be found in the writings of Nicole Oresme, the Renaissance astronomer and mathematician (ca. 1325–1382). Oresme discussed Cicero's example of the number of stars but, writing only a few centuries before the earliest known probability calculations, there is a clear difference:

The number of stars is even; the number of stars is odd. One of these statements is necessary, the other impossible. However, we have doubts as to which is necessary, so that we say of each that it is possible.... The number of stars is a cube. Now indeed, we say that it is possible but not, however, probable or credible or likely [*non tamen probabile aut opinabile aut verisimile*], since such numbers are much fewer than others.... The number of stars is not a cube. We say that it is possible, probable, and likely.... [Oresme 1966, p. 385]

To some, the revolutionary content of this passage lies in its quasinumerical assertion of the improbability of the number of stars being a cube (due to the infrequency of cubic numbers). But its real novelty is Oresme's willingness to extend the realm of the probable. Having made that transition, the frequency-based assertions of probability and improbability he makes follow naturally.

Thus the key step in the birth of mathematical probability – the identification of fundamental probability sets in order to quantify probability – while seemingly so natural, in fact contains a major presupposition. The ancients used symmetry arguments to destroy belief, where we use them to quantify it. This "conceptual revolution" culminated in the 20th century statistical resort to physical randomization (e.g., in sampling, randomized clinical trials, and Monte Carlo simulations): the paradox of deliberately imposing disorder to acquire information. The uses of randomization throughout the ancient and medieval world, in contrast, although common and widespread (for example, in games of chance and fair allocation) all depended, in one way or another, solely on its property of loss of information.

But while the use of symmetry made the calculus of probabilities possible, it also contained the seeds of future confusion.

5. CHANCE AND EQUIPOSSIBILITY

The birth of probability was not an untroubled one. Probabilities are usually classified into two major categories – epistemic and aleatory – and a multitude of subcategories: propensities, frequencies, credibilities, betting odds, and so on. In settings where an FPS exists, all of these will usually have a common value, and the necessity of distinguishing among the different meanings is not a pressing one. But as the initial successes of the "doctrine of chances" spurred on its application to other spheres, this happy state of affairs ceased and the need for distinctions became inevitable.

Just what the proper domains of chance and probability were, however, remained unclear. For the calculus of probabilities was initially the "doctrine of chances," and paradoxically, while the Greeks failed to extend the realm of the probable to include fundamental probability sets, in the early days of the doctrine of chances some thought the notion of chance *only* applicable to such settings. A few examples will suggest the difficulties and confusions that occurred.

1. *Arbuthnot and the sex-ratio.* In 1711, Dr. John Arbuthnot, a Scottish writer, physician to Queen Anne, and close friend of Swift and Pope, published a short paper in the *Philosophical Transactions of the Royal Society*, entitled 'An Argument for Divine Providence Taken From the Constant Regularity Observed in the Births of Both Sexes.' Using statistics from the London Bills of Mortality for the preceding 82 years, Arbuthnot observed that male births had exceeded female births in London for each year from 1629 to 1710. Noting that if male and female births were equally likely, the probability of such an outcome was extremely small (1 in 2^{82}), Arbuthnot rejected the hypothesis of equilikehood, making in effect the earliest known statistical test of significance. But Arbuthnot did not conclude that male and female births possessed unequal probabilities. Instead, he rejected outright the possibility that sex was due to chance, concluding that the excess of males was due to the intervention of divine providence; that "... it is Art, not Chance, that governs" (Arbuthnot 1711, p. 189).

In contrasting art with chance, Dr. Arbuthnot was merely displaying his classical erudition; the dichotomy between *techne* (art) and *tyche* (chance) being a commonplace of Greek philosophy.[13] What *is* new is his belief that chance is only operative when probabilities are equilikely; that otherwise some outside force must be acting, causing the imbalance, and that one could no longer refer to chance. His specific line of reasoning was quickly faulted by Nicholas Bernoulli: if sex is likened to tossing a 35-sided die, with 18 faces labelled "male," and 17 labelled "female," then Arbuthnot's data are entirely consistent with the outcome of chance.[14] This response to Arbuthnot's argument does not dispute that chance is limited to fundamental probability sets; it simply points out that more that one FPS is possible.

Arbuthnot's juxtaposition of chance and cause, and his belief that chances must be equal, is echoed in Hume. For Hume chance "properly speaking, is merely the negation of a cause":

> Since therefore an entire indifference is essential to chance, no one chance can possibly be superior to another, otherwise than as it is compos'd of a superior number of equal chances. For if we affirm that one chance can, after any other manner, be superior to another, we must at the same time affirm, than there is something, which gives it superiority, and determines the event rather to that side that the other: That is, in other words, we must allow of a cause, and destroy the supposition of chance; which we had before establish'd. A perfect and total indifference is essential to chance, and one total indifference can never in itself be either superior or inferior to another. This truth is not peculiar to my system, but is acknowledg'd by every one, that forms calculations concerning chances. [*Hume* 1739, p. 125]

Thus, for Hume, not merely the mathematical calculation of chances but the very existence of chance itself is dependent on an "entire," "perfect," and "total indifference" among the different possibilities. Was this "acknowledg'd by every one?" Examination of the works of Bernoulli, DeMoivre, and Laplace does not entirely bare out this claim. There the equality of chances appears as a mathematical device, not a metaphysical necessity. Nevertheless, the contrast of chance with "art," "design," or "cause," that "something, which gives it superiority," is a recurrent theme. De Moivre suggests that "we may imagine Chance and Design to be, as it were, in Competition with each other" (De Moivre 1756, p. v). "Chance" and "Design" here no longer means the presence and absence of a stochastic element, but a lack of uniformity in the probability distribution. Answering Nicholas Bernoulli, De Moivre says yes, Arbuthnot's birth data is consistent with an 18:17 ratio, but "this Ratio once discovered, and *manifestly serving to a wise purpose*, we conclude the Ratio itself, or if you will the *Form of the Die*, to be an Effect of *Intelligence and Design*" (De Moivre 1756, p. 253).

Uniformity in distribution was to be increasingly equated with absence of design or law, departure from uniformity with their presence. A famous example is Michell's argument in 1767 that optically double or multiple stars were physically so. Michell calculated that the observed clustering of stars in the heavens exceeded what could reasonably be expected if the stars were distributed at random (i.e., uniformly) over the celestial sphere, inferring "either design, or some general law" due to "the greatness of the odds against things having been in the present situation, if it was not owing to some such cause" (Michell 1767, p. 243). Michell's argument was the focus of debate for a brief period during the middle of the 19th century, a key issue being precisely this equation of uniformity with absence of law.[15]

The elements of a fundamental probability set enjoy this status for reasons which are both *aleatory* (i.e., physical or objective) and *epistemic*. The dichotomy between chance and design involves primarily the aleatory aspect of the FPS. Throughout the 18th century, the elements of an FPS were often defined in terms of equipossibility, a terminology which, as Hacking notes (1975, Chapter 14), permitted a blurring of the aleatory and epistemic aspects. The literature of the period furnishes many instances of this duality. In the *Ars Conjectandi*, for example, James Bernoulli refers to cases which are "equally possible, that is to say, each can come about as easily as any other" (*omnes casus aeque possibiles esse, seu pari facilitate evenire posse*). Laplace, on the other hand, in his *Essai philosophique*, states the following famous – and purely epistemic – criterion:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence.... [Laplace 1952, p. 6]
>
> If [the various cases] are not [equally possible], we will determine first their respective possibilities, whose exact appreciation is one of the most delicate points of the theory of chance. [Laplace 1952, p. 11]

To assign equal probability to cases "such as we may be equally undecided about" is the notorious *principle of insufficient reason*. Although Laplace did not view it as controversial, many in the 19th century did. What determines when cases are equally probable, possible, or likely? This epistemological ambiguity in the meaning and determination of an FPS led inevitably to controversy in its application.

2. *D'Alembert and De Morgan.* For example, what is the chance of getting at least one head in two tosses of a fair coin? The standard solution to this problem regards the four possible outcomes of tossing a coin twice – HH, HT, TH, TT – as equally likely; since 3 out of these four cases are favorable, the probability is 3/4. In 1754, however, the French *philosophe* Jean Le Rond D'Alembert (1717–1783) advanced a different solution in his article 'Croix ou pile' in the *Encyclopedie*. D'Alembert reasoned that one would stop tossing the coin as soon as the desired head came up, so that there are really only three possible outcomes – H, TH, TT – two of which are favorable, and hence the probability is 2/3.

D'Alembert was far from being the first distinguished mathematician to make an elementary error of this type, but he is perhaps unique in the doggedness with which he subsequently defended his answer. Indeed, this was only the first of several instances where D'Alembert was led to disagree with the standard answers of the calculus of probabilities, and "with this article, the renowned mathematician opened a distinguished career of confusion over the theory of probabilities" (Baker 1975, p. 172).[16]

D'Alembert's criticisms were largely greeted with scorn and ridicule, but seldom seriously discussed. Laplace, for example, remarks that the probability would indeed be 2/3 "if we should consider with

D'Alembert these three cases as equally possible . . ." (1952, p. 12), but he limits himself to giving the standard calculation without explaining why one set of equipossible cases is preferable to another.

The D'Alembert fallacy is possible because of the ambiguity in the concept of equipossibility and the Laplacean definition of probability. Laplace's treatment of these questions, although not confused, fails to come to grips with the fundamental issues. For one of the few serious discussions of D'Alembert's argument, one must turn to the writings of Augustus De Morgan, Laplace's most enthusiastic and influential English expositor during the first half of the 19th century.

De Morgan argued that there are essentially two very distinct considerations involved in the assessment of numerical probabilities. The first of these is *psychological*: the measurement and comparison of "the impressions made on our minds by different prospects," as in a judgment of equiprobability among alternatives. The second is *mathematical*: the rational use of such measures or comparisons, as in the computation of the probability of a complex event involving simpler, equiprobable outcomes. The two questions differ in that "any given answer to the first may admit of dispute," while "there is no fear of mathematics failing us in the second," (De Morgan 1845, p. 395).

Armed with this distinction, De Morgan was able to analyze the D'Alembert fallacy:

> [W]ith regard to the objection of D'Alembert . . . , we must observe that if any individual really feel himself certain, in spite of authority and principle, as here laid down, that the preceding cases are equally probable, he is fully justified in adopting 2/3 instead of 3/4, till he see reason to the contrary, which it is hundreds to one he would find, if he continued playing for a stake throughout a whole morning, that is, accepting bets of two to one that H would not come up once in two throws, instead of requiring three to one. . . . The individual just supposed, has applied correct mathematics to a manner in which he feels obliged to view the subject, in which we think him wrong, but the error is in the first of the two considerations [above], and not in the second. [De Morgan 1845, p. 401]

Despite its interest, De Morgan's discussion is ultimately unsatisfactory. The choice of an FPS is described as a psychological consideration (which would suggest a subjectivist viewpoint), but the phrase "in which we think him wrong" suggests an objectivistic one. De Morgan appeals to experience to justify the classical choice of FPS in the D'Alembert problem, although probabilities for De Morgan were degrees of belief rather than empirical frequencies. The Laplacean view of probability was one of rational degree-of-belief, but his followers were understandably reluctant to uncouple probability from frequencies although, not surprisingly, unable to provide a logical description of the choice of FPS.

De Morgan later returned to the D'Alembert example in his *Formal Logic* (1847, pp. 199–200), and his brief discussion there is also interesting:

> [I]t may happen that the state of mind which *is*, is not the state of mind which should be. D'Alembert believed that it was *two* to *one* that the first head which the throw of a halfpenny was to give would occur before the third throw; a juster view of the mode of applying the theory would have taught him it was *three* to *one*. But he *believed* it, and thought he could show reason for his belief: to him the probability *was* two to one. But I shall say, for all that, that the probability *is* three to one: meaning, that in the universal opinion of those who examine the subject, the state of mind to which a person *ought* to be able to bring himself is to look three times as confidently upon the arrival as upon the non-arrival.

When De Morgan says that, for D'Alembert, "the probability *was*," the word probability is being used in a psychological or personalist sense; when he says "the probability *is*," the sense is logical or credibilist. But to say that the probability is three to one because that is "the universal opinion of those who examine the subject," while certainly candid, is hardly a devastating refutation of D'Alembert.

De Morgan deserves considerable credit for distinguishing between the psychological process of identifying a set of outcomes as equipossible, and the mathematical use of such a set to calculate probabilities, as well as his (implicit) distinction between the subjective and objective senses of probability. Where he fails is in his account of why the probability "is" three to one, and what empirical justification, if any, such a statement requires. These, however, were basic questions for which the theory of his day had no answer.

In the later half of the 19th century, a serious attack was mounted on epistemic probability and the principle of insufficient reason, and a direct confrontation with such questions could no longer be avoided.

6. THE PRINCIPLE OF INSUFFICIENT REASON

The contributions of Laplace represent a turning point in the history of probability. Before his work, the mathematical theory was (with the exception of the limit theorems of Bernoulli and DeMoivre) relatively unsophisticated, in effect a subbranch of combinatorics; its serious applications largely confined to games of chance and annuities. All this changed with Laplace. Not only did he vastly enrich the mathematical theory of the subject, both in the depth of its results and the range of the technical tools it employed, he demonstrated it to be a powerful instrument having a wide variety of applications in the physical and social sciences. Central to his system however, was the use of the so-called principle of insufficient reason.[17]

The 19th century debate about the validity of the principle of insufficient reason involved, of necessity, much broader issues. Is probability empirical or epistemic in nature? Can a probability be meaningfully assigned to any event? Are all probabilities numerically quantifiable? Beginning in the 1840's, and continuing on into the 20th century, a number of eminent British mathematicians, philosophers, and scientists began to address such questions, including De Morgan, Ellis, Mill, Forbes, Donkin, Boole, Venn, Jevons, MacColl, Edgeworth, Keynes, Ramsey, Jeffreys, and Broad.

1. *Donkin.* A comprehensive discussion of this literature would be beyond the scope of the present paper. Instead, we will confine our attention primarily to the contributions of William Fishburn Donkin, Savilian Professor of Astronomy in the University of Oxford from 1842 to 1869. Donkin wrote two papers on mathematical probability. One of these concerned the justification for the method of least squares and, although a valuable contribution to that subject, will not concern us here. The other paper is modestly titled 'On Certain Questions Relating to the Theory of Probabilities' (Donkin 1851). Donkin's paper, although little-known, is a lucid and careful attempt to clarify the foundations of the subject. It was written in response to criticisms by Forbes and others of Michell's argument that stars that are optically double are also physically so.

Donkin begins by stating that

> It will, I suppose, be generally admitted, and has often been more or less explicitly stated, that the subject matter of calculation in the mathematical theory of probabilities is *quantity of belief.*

There were some dissenters to this view of probability at the time Donkin wrote (e.g., Ellis 1844; Mill 1843), but they were few in number and, due at least in part to the influence of De Morgan, Laplace's views held sway in England.[18]

Donkin's philosophical view of probability may be summarized as *relative, logical, numerical, and consistent.* Probability is relative in the sense that it is never "inherent in the hypothesis to which it refers," but "always *relative* to a state of knowledge or ignorance." Nevertheless, Donkin was not a subjectivist, because he also believed probability to be

> *absolute* in the sense of not being relative to any individual mind; since, the same information being presupposed, all minds *ought* to distribute their belief in the same way.

Ultimately, any such theory of logical probability must resort to the principle of insufficient reason, and Donkin's was no exception. Indeed, if anything he saw its role as even more central to the theory than did Laplace:

> ... the law which must always be made the foundation of the whole theory is the following: – *When several hypotheses are presented to our mind, which we believe to be mutually exclusive and exhaustive, but about which we know nothing further, we distribute our belief equally amonst them.*

Although Boole's detailed and influential criticism of the appeal to insufficient reason was still several years off (Boole 1854, pp. 363–375), Robert Leslie Ellis had already attacked its use on the grounds that it "erected belief upon ignorance" (Ellis 1850, p. 325). Donkin's response was to stake out a limited claim for the theory:

> [The force of] the argument commonly called the "sufficient reason" ... in all cases depends (as it seems to me) upon a previous *assumption* that *an intelligible law exists* concerning the matter in question. If this assumption be admitted, and if it can be shown that there is only *one* intelligible law, then that must be the actual law.... A person who should dispute the propriety of dividing our belief equally amongst hypotheses about which we are equally ignorant, ought to be refuted by asking him to state *which is to be preferred*. He must either admit the proposed law, or maintain that there is no law at all.

This observation would not have disarmed Ellis, Boole, or Venn, who indeed denied the existence of any (determinate in the case of Boole) law at all. But it did draw the line clearly. Its vulnerability, as Boole realized, is simply that two or more sets of "mutually exclusive and exhaustive" hypotheses may present themselves "about which we know nothing further," and which give rise to incompatible probability assignments. Ramsey saw it as a virtue of the subjectivistic theory that it eluded this dilemma by dispensing with the requirement of a unique law, admitting more than one probability assignment as possible (Ramsey 1926, pp. 189–190).

But can one calculate probabilities no matter how complex the setting or information available? Cournot, for example, had earlier argued that there were three distinct categories of probability – objective, subjective, and philosophical, the last involving situations whose complexity precluded mathematical measurement.[19]

Donkin thought such arguments, essentially pragmatic in nature, not to the point:

> . . . I do not see on what ground it can be doubted that every definite state of belief concerning a proposed hypothesis is in itself capable of being represented by a numerical expression, however difficult or impracticable it may be to ascertain its actual value. . . . [It is important to distinguish] the difficulty of *ascertaining numbers* in certain cases from a supposed difficulty of *expression by means of numbers*. The former difficulty is real, but merely relative to our knowledge and skill; the latter, if real, would be absolute, and inherent in the subject matter, which I conceive not to be the case.

This was an important distinction. It expresses a tenet of faith of logical probability: that all probabilities can, in principle be measured. On a basic philosophical level, such theories have never really answered Ramsey's simple criticism:

> It is true that about some particular cases there is agreement, but these somehow paradoxically are always immensely complicated; we all agree that the probability of a coin coming down heads is 1/2, but we can none of us say exactly what is the evidence which forms the other term for the probability relation about which we are then judging. If, on the other hand, we take the simplest possible pairs of propositions such as 'This is red', and 'That is blue', or 'This is red' and 'That is red', whose logical relations should surely be easiest to see, no one, I think, pretends to be sure what is the probability relation between them. [Ramsey 1926]

2. *Boole.* The first influential critic of the principle of insufficient reason was Boole. He says of its derivation:

> It has been said, that the principle involved in the above and in similar applications is that of the equal distribution of our knowledge, or rather of our ignorance – the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability. I apprehend, however, that this is an arbitrary method of procedure. [Boule 1854, p. 370]

As we have seen earlier (Section 3), to justify his criticism Boole pointed to instances where it was possible to partition the sample space of possible outcomes in different ways, each of which could plausibly

be viewed as equipossible. Boole's criticisms, unfortunately, became more confusing as he attempted to clarify them. One might be forgiven, for example, for interpreting the passage just quoted as a clear rejection of the principle. But Boole later wrote:

> I take this opportunity of explaining a passage in the *Laws of Thought*, p. 370, relating to certain applications of the principle. Valid objection lies not against the principle itself, but against its application through arbitrary hypotheses, coupled with the assumption that any result thus obtained is necessarily the true one. The application of the principle employed in the text and founded upon the general theorem of development in Logic, I hold to be *not* arbitrary. [Boole 1862]

Perusal of "the application of the principle employed in the text" reveals it to be of the balls in an urn type, and what Boole now appears to be defending might be called the principle of *cogent* reason: if one possesses some information about the different alternatives, but this information is equally distributed amongst them, then one is justified in assigning the alternatives equal probability.

Boole appeared to regard both probabilistic independence (which he used extensively in his system) and uniformity of distribution as assumptions of *neutrality*, in each case a *via media* between conflicting extremes. There is a simple geometric sense in which this is true for the assumption of uniformity: the uniform distribution on $n + 1$ elements is the barycenter of the n-dimensional simplex of all probability distributions. But once more the consequences of a symmetry assumption lurk only partially visible. For depending on the use being made of a probability distribution, symmetrical or uniform distributions can often represent an extreme type of behavior. A good example of this involves the "birthday paradox": in a group of 23 or more people, the odds exceed 1/2 that at least two persons share a birthday in common (Feller 1968, p. 33). The calculation on which this statement is based assumes that births occur uniformly throughout the year. Although empirically false (see, e.g., Izenman and Zabell 1982), this does not affect the validity of the conclusion: the probability of a birthday "match" is minimized when the distribution of births is uniform (so that the probability of a match will be even greater under the true distribution).

It is difficult to assess Boole's immediate impact on his contemporaries. As the distinguished author of *The Laws of Thought*, his views on probability were certainly treated with respect. Nevertheless, they were highly idiosyncratic and confused in important respects.[20] Given the complexity and unattractiveness of his own system, and lacking the alternative philosophical foundation to the Laplacean edifice that was later provided by Venn's *Logic of Chance*, there was an obvious reluctance to abandon the classical theory. Nevertheless, his pointing to the fundamental ambiguity in the principle of insufficient reason was a lasting contribution, remembered long after the rest of his work on probability was forgotten.

Donkin represents what may be the highwater mark in the defense of the Laplacean position; Boole was its first influential English critic. After Boole and Venn the Laplaceans were on the defensive, first in the philosophical, later in the statistical and scientific communities. In response to the criticisms of Boole and his successors, many attempts were made to state unambiguous formulations of the principle of insufficient reason (e.g., by von Kries and Keynes), but their increasing obscurity and complexity ensured their rejection.[21]

The debate about the principle of insufficient reason and its consequence, Laplace's rule of succession, tapered off in the 1920s. This was partly because Ramsey's 1926 essay 'Truth and Probability' made the principle superfluous as a foundation for epistemic probability. When Fisher and Neyman produced statistical methodologies independent of the Bayes–Laplace edifice, Bayesian statistics essentially disappeared, only to be resuscitated by Savage nearly a quarter of a century later with the publication in 1954 of his *Foundations of Statistics*.

Savage's conversion to subjectivism occurred after he became acquainted with de Finetti's work, and his writings were largely responsible for bringing it into the mainstream of philosophical and statistical thought. At the center of de Finetti's system was the notion of exchangeability, and thus, initially exorcised, symmetry re-entered epistemic probability.

7. WHAT IS TO BE DONE?

Symmetry arguments are tools of great power; therein lies not only their utility and attraction, but also their potential treachery. When they are invoked one may find, as did the sorcerer's apprentice, that the results somewhat exceed one's expectations. Nevertheless, symmetry arguments enjoy an honored and permanent place in the arsenal

of probability. They underlie the classical definition of probability that held sway for over two centuries, are central to virtually all quantitative theories of induction, appear as exchangeability assumptions in subjectivist theories, and, in the guise of group-invariance, still play an important role in modern theoretical statistics. Their use calls for judicious caution rather than benign neglect.

The ambiguity underlying the proper role of symmetry assumptions in the theory of probability stems in part from a corresponding ambiguity about the role the axioms play in the various axiomatic formulations of probability. Do the axioms enjoy a privileged status *vis-à-vis* their deducible consequences? Are they supposed to be intuitively more evident or simpler in form? If the justification for the axioms is their intuitive acceptability, what if some of their consequences violate those intuitions? As in so many cases, one can identify two polar positions on such issues, that of the *left-wing dadaists* and the *right-wing totalitarians*.[22]

The left-wing dadaists not only demand that the axioms be grounded in our intuitions, but that *all* deducible consequences of the axioms must be intuitively acceptable as well. Intuitive acceptability was the warrant for the axioms in the first place, and since there is no obvious reason to favor certain intuitions over others, all must be satisfied. If the consequences of a set of axioms violate our intuitions, then those axioms must be abandoned and replaced. A leading exponent of this position is L. Jonathan Cohen.[23]

The problem with such a position is that our intuitions, or at least our *untutored* intuitions, are often mutually inconsistent and any consistent theory will necessarily have to contradict some of them. During the last two decades many psychologists, notably Daniel Kahneman and Amos Tversky, have demonstrated that popular intuitions are often inconsistent not merely with the standard axioms of probability, but with essentially *any* possible axiomatization of probability; that "people systematically violate principles of rational decision-making when judging probabilities, making predictions, or otherwise attempting to cope with probabilistic tasks" (Slovic, Fischhoff, and Lichtenstein 1976).[24]

The right-wing totalitarians, on the other hand, believe that once an axiom system is adopted, one must accept without question every consequence that flows from it. One searches within one's heart, discovers the basic properties of belief and inference, christens them axioms, and then all else follows as logical consequence. Once the axioms are brought to the attention of unbelievers, they must, like Saul on the road to Damascus, be smitten by instantaneous conversion or they stand convicted of irrational obtuseness. One representative of this position is E. T. Jaynes, who dates his adherence to Bayesianism to the time when he encountered Cox's axiomatization of epistemic probability, and who views the Shannon axioms for entropy as an unanswerable foundation for his method of maximum entropy.[25]

This position errs in giving the axioms too distinguished a position, just as the previous position gave them too little. A set of axioms A, together with $T(A)$, the theorems deducible from it, forms a self-consistent whole S. Let us say that any subset $B \subseteq S$, such that $B \cup T(B) = S$, is an *axiom-system for S*. Mathematically speaking, all possible axiom-systems for S must be regarded as starting out on an equal footing, and which axiom-system is ultimately chosen is essentially a matter of preference, depending on considerations such as simplicity, elegance, and intuitive acceptability.

The key point is that having tentatively adopted an axiom system, one is not obligated to uncritically accept its consequences. In both formal logic and subjective probability, the theory polices sets of beliefs by testing them for inconsistencies, but it does not dictate how detected inconsistencies should be removed. If, as was the case with some of the symmetry assumptions previously discussed, the consequences are deemed unacceptable, then the assumption will be discarded. If, on the other hand, the axioms seem compelling, as in mathematical probability, then surprising consequences such as the birthday paradox will be regarded as valuable correctives to our erroneous, untutored intuitions; that is why the theory is useful. What is or should be at play is a dynamic balance. As Nelson Goodman argues:

> Inferences are justified by their conformity to valid general rules, and ... general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend*. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either [Goodman 1979, p. 64].

Symmetry assumptions must therefore be tested in terms of the

particular inferences they give rise to. But – and this is the rub – particular inferences can only be reasonably judged in terms of particular situations, whereas symmetry assumptions are often proposed in abstract and theoretical settings devoid of concrete specifics.[26]

Fundamentally at issue here are two very different approaches to the formulation of a logic of probability. Extreme subjectivists adopt a *laissez faire* approach to probability assignments, emphasizing the unique aspects attending the case at hand. They do not deny the utility of symmetry arguments, but, as Savage remarks, they "typically do not find the contexts in which such agreement obtains sufficiently definable to admit of expression in a postulate" (Savage 1954, p. 66). Such arguments fall instead under the rubric of what I. J. Good terms "suggestions for using the theory, these suggestions belonging to the technique rather than the theory" itself (Good 1952, p. 107).

Proponents of logical theories, in contrast, believe (at least in principle) that if the evidence at one's disposal is stated with sufficient precision in a sufficiently rich language then agreement can be forced via considerations of symmetry. At the level of ordinary language such claims founder at the very outset on Ramsey's simple objection (quoted earlier in Section 6). Instead, simple model languages are introduced and probabilities computed "given" statements descriptive of our state of knowledge. Such formal systems do not escape subjectivism, they enshrine it in the equiprobable partitions assumed.

Practical attempts to apply logical probability always seem to lead back to discussions about events "concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it." Such attempts are ultimately divorced from reality, if only because understanding the very meaning of the words employed in describing an event already implies knowledge about it. Thus, it is not surprising that the three leading 20th century proponents of logical probability – Keynes, Jeffreys, and Carnap – all eventually recanted to some extent or another.[27] Carnap, for example, wrote

> I think there need not be a controversy between the objectivist point of view and the subjectivist or personalist point of view. Both have a legitimate place in the context of our work, that is, the construction of a system of rules for determining probability values with respect to possible evidence. At each step in the construction, a choice is to be made; the choice is not completely free but is restricted by certain boundaries. Basically, there is merely a difference in attitude or emphasis between the subjectivist tendency to emphasize the existing freedom of choice, and the objectivist tendency to stress the existence of limitations. [Carnap 1980, p. 119]

This little-known, postumously published passage is a substantial retreat from the hard-core credibilism of the *Logical Foundations of Probability*. But it was inevitable. Symmetry arguments lie at the heart of probability. But they are tools, not axioms, always to be applied with care to specific instances rather than general propositions.

8. ENVOI

As a final illustration of the seductive nature of symmetry arguments in probability, and as a challenge to the reader, I end with a little puzzle, which I will call the *exchange paradox*:[28]

> A, B, and C play the following game. C acts as referee and places an unspecified amount of money x in one envelope and amount $2x$ in another envelope. One of the two envelopes is then handed to A, the other to B.
>
> A opens his envelope and sees that there is \$10 in it. He then reasons as follows: "There is a 50–50 chance that B's envelope contains the lesser amount x (which would therefore be \$5), and a 50–50 chance that B's envelope contains the greater amount $2x$ (which would therefore be \$20). If I exchange envelopes, my expected holdings will be (1/2)\$5 + (1/2)\$20 = \$12.50, \$2.50 in excess of my present holdings. Therefore I should try to exchange envelopes."
>
> When A offers to exchange envelopes, B readily agrees, since B has already reasoned in similar fashion.

It seems unreasonable that the exchange be favorable to both, yet it appears hard to fault the logic of either. I will resist the temptation to explain what I take to be the resolution of the paradox, other than noting that all hinges on A's apparently harmless symmetry assumption that it is equally likely that B holds the envelope with the greater or the lesser amount.[29]

Department of Mathematics
Northwestern University

NOTES

[1] The symbol $\binom{n}{k}$ denotes the binomial coefficient $n!/[k!(n-k)!]$. Note that in the theorem the sequence is assumed to be infinite; this requirement is sometimes overlooked, although it is necessary for the general validity of the theorem.

[2] There also exist finite forms of de Finetti's theorem, which permit one to dispense with the assumption that the number of trials is infinite. In such cases the integral mixture is either replaced by a discrete sum or serves as an approximation to the exact probability; see Diaconis and Freedman (1980a).

[3] The existence of limiting frequencies for infinite exchangeable sequences follows from their stationarity, and is an immediate consequence of the ergodic theorem; see, e.g., Breiman (1968, p. 118, Theorem 6.28).

[4] For further discussion of Bayes's scholium, see Murray (1930), Edwards (1978). For an interesting account of how Bayes's argument has often been misconstrued by statisticians to fit their foundational preconceptions, see Stigler (1982).

[5] It is an interesting historical footnote that Johnson's derivation almost never appeared. After the appearance of the third volume of his *Logic* in 1924, Johnson began work on a fourth volume, to be devoted to probability. Unfortunately, Johnson suffered a stroke in 1927, and the projected work was never finished. Drafts of the first three chapters were edited by R. B. Braithwaite and published posthumously as three separate papers in *Mind* during 1932. Johnson's mathematical derivation of the continuum of inductive methods from the sufficientness postulate appeared as an appendix in the last of the three. G. E. Moore, then editor of *Mind*, questioned whether so technical a result would be of general interest to its readership, and it was only on the insistence of Braithwaite that the appendix was published (Braithwaite 1982, personal communication).

[6] For further information about Johnson's sufficientness postulate, and a complete version of his proof, see Zabell (1982).

[7] In brief, this is because even when one lacks information about specific, identifiable categories, one may possess information about the vector of *ordered* probabilities. (For example, one may know that a die is biased in favor of one face, but not know which face it is.)

[8] See generally Heath (1913, Chapter 4); Kahn (1960); Dicks (1970, Chapter 3). For the original Greek texts of the fragments of Anaximander, with accompanying English translation, commentary, and discussion, see Kirk and Raven (1957, Chapter 3).

[9] Perhaps the most pessimistic assessment of the state of our information is that of Dicks (1970, pp. 45–46).

[10] In its general form (neither of two exactly symmetrical alternatives will occur), it also crops up from time to time in 19th century philosophical discussions of probability. Two examples are (1) Bolzano: "... if we are to have a rational expectation that a certain result will take place, for example that Caius will draw a certain ball from several balls in an urn, then we must presuppose that the relation between these balls and Caius is such that the reasons for drawing that particular ball are not exactly like the reasons for drawing some other ball, since otherwise he wouldn't draw any" (Bolzano 1837, p. 245 of 1972 edition.); (2) Cook Wilson: "... if a number of cases, mutually exclusive ..., were in the nature of things equally possible, not one of them could happen. If the claim of any one of them in reality were satisfied, so must the claim of any other, since these claims are equal, and therefore if one happens all must, but by hypothesis if one happens no other can; thus the only possible alternative is that none of them can happen" (Wilson 1900, p. 155).

[11] Aristocles, quoted in Long (1974, p. 81); c.f. Diogenes Laertius, *Life of Pyrrho* 9.107; Sextus Empiricus, *Outlines of Pyrrhonism* 1.8. For general information on the Pyrrhonian skeptics, see Stough (1969, Chapter 2); Long (1974, pp. 75–88). The *ou mallon* argument itself is discussed in some detail by DeLacy (1958).
[12] See generally Stough (1969, pp. 50–66); Long (1974, pp. 95–99).
[13] See, e.g., Plato, *Laws* 709, 889 b–d; Aristotle, *Metaphysics* 1070ab. (Strictly speaking, Aristotle distinguishes between *automaton* (chance, spontaneity) and *tyche* (luck, fortune).
[14] For further discussion of Arbuthnot, see Hacking (1965, pp. 75–77); Hacking (1975, Chapter 18); Pearson (1978, pp. 127–133, 161–162).
[15] For a recent and very readable account of the dispute, see Gower (1982). Similar issues arose in later discussions of geometrical probability: what does it mean to select points (or lines, or triangles) at random? Venn (1888, pp. 100–101), reporting one such discussion, quotes the English mathematician Crofton as asserting that "at random" has "a very clear and definite meaning; one which cannot be better conveyed than by Mr Wilson's definition, 'according to no law'...." "Mr. Crofton holds," Venn continues, "that any kind of *unequal* distribution [of points in a plane] would imply law," to which Venn retorts, "Surely if they tend to become *equally* dense this is just as much a case of regularity or law." Where James Bernoulli had attempted to subsume the probability of causes under that of chances (to use Hume's terminology), the frequentist Venn subsumes the probability of chances under that of causes.
[16] See generally Todhunter (1865, Chapter 13); Baker (1975, pp. 171–180); Pearson (1978, Chapter 12). For a recent but unconvincing attempt at rehabilitation, see Daston (1979).
[17] Laplace nowhere actually uses this term, which is of later origin. Writing in 1862, Boole refers to "that principle, more easily conceived than explained, which has been differently expressed as the 'principle of non-sufficient reason', the principle of equal distribution of knowledge or ignorance' [footnote omitted], and the 'principle of order'," (Boole 1862).
[18] When Donkin wrote his paper the first frequentist theories (apart from occasional allusions in the earlier literature) were less than a decade old. As Porter (1986, p. 77) notes, "in 1842 and 1843, four writers from three countries independently proposed interpretations of probability that were fundamentally frequentist in character." These four – Jakob Friedrick Fries in Germany, Antoine Augustin Cournot in France, and Richard Leslie Ellis and John Stuart Mill in England – were the harbingers of an increasingly empirical approach to probability. (Curiously, after correspondence with the astronomer John Herschel, Mill actually withdrew his objections to Laplace's epistemic view of probability from the second (1846) and later editions of his *Logic*; see Strong (1978).) Despite this early efflorescence, the frequency theory did not begin to gain widespread acceptance until its careful elaboration, nearly a quarter of a century later, in John Venn's *Logic of Chance* (1st ed. 1866). For discussion of the work of Fries, Cournot, Ellis, and Mill, see Porter (1986, pp. 77–88), Stigler (1986, pp. 195–200); for discussion of Venn's *Logic*, Salmon (1980).
[19] The argument that some probabilities are "philosophical" (i.e., inherently non-numerical) was often made by those who thought the mathematical theory had out-reached its grasp. Strong (1976, p. 207, n. 5) notes the use of the distinction in K. H. Frömmichen's 1773 work, *Über die Lehre des Wahrscheinlich*, "the earliest ... that I have been able definitely to date," as well the better known treatment in Kant's *Logik* of 1781. See von Wright (1957, p. 217, n. 9) for further references to the 19th century literature. In addition to the names given there, one could add those of the Scottish philosopher Dugald Stewart and the English jurists Starkie, Wills, and Best. For the related criticisms of the French positivists Destutt de Tracy, Poinsot, and Comte, see Porter (1986, p. 155) and Stigler (1986, pp. 194–195).

[20] Many of these are touched on by Keynes in scattered passages throughout his *Treatise on Probability* (1921). Hailperin (1976) is a useful attempt at rational reconstruction. For discussion of Boole's criticism of the Laplace/De Morgan analysis of inductive reasoning in terms of probability, see the excellent article of Strong (1976).
[21] See generally Keynes (1921, Chapters 4 and 6).
[22] There is obviously an element of intentional caricature in what follows, although perhaps less than might be supposed.
[23] "... ordinary human reasoning ... cannot be held to be faultily programmed: it sets its own standards" (Cohen 1981, p. 317).
[24] Much of this work is summarized in Kahneman, Slovic, and Tversky (1982).
[25] Although not readily available, Jaynes's early Socony Mobil Oil lecture notes (Jaynes 1958) provide a vigorous and very readable exposition of his viewpoint.
[26] There are some notable exceptions to this. W. E. Johnson, for example, in discussing his sufficientness postulate, argued that:

"the postulate adopted in a controversial kind of theorem cannot be generalized to cover all sorts of working problems; so it is the logician's business, having once formulated a specific postulate, to indicate very carefully the factual and epistemic conditions under which it has practical value." (Johnson 1932, pp. 418–419)

[27] For Keynes's recantation, see Good (1965, p. 7). In the third edition of his book *Scientific Inference*, Jeffreys suggests that in controversial cases the appropriate choice of reference prior could be decided by an international panel of experts. Such a position is obviously incompatible with credibilism as usually understood. For Carnap, see the text *infra*.
[28] I first heard the paradox from Steve Budrys of the Odesta Corporation, on an otherwise unmemorable night at the now defunct Chessmates in Evanston. It does not originate with him, but I have been unable to trace its ultimate source.
Note added in proof: Persi Diaconis and Martin Gardner inform me that the paradox is apparently due to the French mathematician Maurice Kraitchik; see Maurice Kraitchik, *Mathematical Recreations*, 2nd ed. (New York: Dover, 1953), pp. 133–134. In Kraitchik's version two persons compare their neckties, the person with the less valuable necktie to receive both.
[29] I thank Persi Diacanis, David Malament and Brian Skyrms for helpful comments.

BIBLIOGRAPHY

Arbuthnot, John (1711) 'An argument for divine providence taken from the constant regularity observed in the births of both sexes', *Philosophical Transactions of the Royal Society of London* **27**, 186–190.
Baker, Keith Michael (1975) *Condorcet: From Natural Philosophy to Social Mathematics* (Chicago: University of Chicago Press).
Bayes, Thomas (1764) 'An essay towards solving a problem in the doctrine of chances', *Philosophical Transactions of the Royal Society of London* **53**, 370–418.

Bolzano, Bernard (1837) *Wissenschaftslehre*. Translated 1972 under the title *Theory of Science* (R. George, ed. and trans.) (Berkeley and Los Angeles: University of California Press).
Boole, George (1854) *An Investigation of the Laws of Thought* (London: Macmillan.) (Reprinted 1958, New York: Dover Publications.)
Boole, George (1862) 'On the theory of probabilities', *Philosophical Transactions of the Royal Society of London* **152**, 386–424.
Breiman, Leo (1968) *Probability* (Reading, Mass.: Addison-Wesley).
Broad, C. D. (1975) *Leibniz: An Introduction* (Cambridge University Press).
Carnap, Rudolph (1950) *Logical Foundations of Probability* (The University of Chicago Press. Second edition, 1960).
Carnap, Rudolph (1952) *The Continuum of Inductive Methods* (Chicago: University of Chicago Press).
Carnap, Rudolph (1980) 'A basic system of inductive logic, part II', in *Studies in Inductive Logic and Probability*, volume II (Richard C. Jeffrey, ed.) (Berkeley and Los Angeles: University of California Press) pp. 7–155.
Cohen, L. Jonathan (1981) 'Can human irrationality be experimentally demonstrated?', *The Behavioral and Brain Sciences* **4**, 317–370 (with discussion).
Cournot, Antoine Augustin (1843) *Exposition de la théorie des chances et des probabilités* (Paris: Libraire de L. Hachette).
Daston, Lorraine J. (1979) 'D'Alembert's critique of probability theory', *Historia Mathematica* **6**, 259–279.
De Finetti, Bruno (1937) 'La prévision: ses lois logiques, ses sources subjectives', *Annales de l'Institut Henri Poincaré* **7**, 1–68.
DeLacy, Phillip (1958) 'Ou mallon and the antecedents of ancient scepticism', *Phronesis* **3**, 59–71.
De Moivre, Abraham (1756) *The Doctrine of Chances* (3rd ed.), London.
De Morgan, Augustus (1845) 'Theory of probabilities', *Encyclopedia Metropolitana, Vol. 2: Pure Mathematics* (London: B. Fellowes *et al.*) pp. 393–490.
De Morgan, Augustus (1847) *Formal Logic: Or the Calculus of Inference Necessary and Probable*. London.
Diaconis, Persi (1977) 'Finite forms of de Finetti's theorem on exchangeability', *Synthese* **36**, 271–281.
Diaconis, Persi and Freedman, David (1980a) 'Finite exchangeable sequences', *Annals of Probability* **8**, 745–764.
Diaconis, Persi and Freedman, David (1980b) 'De Finetti's theorem for Markov chains', *Annals of Probability* **8**, 115–130.
Diaconis, Persi and Freedman, David (1980c) 'De Finetti's generalizations of exchangeability', in *Studies in Inductive Logic and Probability*, volume II (Richard C. Jeffrey, ed.) (Berkeley and Los Angeles: University of California Press) pp. 233–249.
Diaconis, Persi and Freedman, David (1985) 'Partial exchangeability and sufficiency', *Statistics: Applications and New Directions*. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (Calcutta: Indian Statistical Institute) pp. 205–236.
Dicks, D. R. (1970) *Early Greek Astronomy to Aristotle* (Ithaca: Cornell University Press).
Donkin, William Fishburn (1851) 'On certain questions relating to the theory of probabilities', *Philosophical Magazine* **1**, 353–368; **2**, 55–60.

Edwards, A. W. F. (1978) 'Commentary on the arguments of Thomas Bayes', *Scandinavian Journal of Statistics* **5**, 116–118.
Ellis, Richard Leslie (1844) 'On the foundations of the theory of probabilities', *Transactions of the Cambridge Philosophical Society* **8**, 1–6.
Ellis, Richard Leslie (1850) 'Remarks on an alleged proof of the "method of least squares" contained in a late number of the *Edinburgh Review*', *Philosophical Magazine* **37**, 321–328.
Feller, William (1968) *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed. (New York: Wiley).
Good, Irving John (1952) 'Rational decisions', *Journal of the Royal Statistical Society B* **14**, 107–114.
Good, Irving John (1965) *The Estimation of Probabilities: An Essay on Modern Bayesian Methods* (Cambridge, Mass.: M. I. T. Press).
Goodman, Nelson (1979) *Fact, Fiction, and Forecast* (3rd ed.) (Indianapolis: Hackett Publishing Company).
Gower, Barry (1982) 'Astronomy and probability: Forbes versus Michell on the distribution of the stars', *Annals of Science* **39**, 145–160.
Hacking, Ian (1965) *Logic of Statistical Inference* (Cambridge University Press).
Hacking, Ian (1967) 'Slightly more realistic personal probability', *Philosophy of Science* **34**, 311–325.
Hacking, Ian (1975) *The Emergence of Probability* (Cambridge University Press).
Hailperin, Theodore (1976) *Boole's Logic and Probability*. Studies in Logic and the Foundations of Mathematics, volume 85 (Amsterdam: North-Holland).
Heath, Sir Thomas (1913) *Aristarchus of Samos: The Ancient Copernicus* (Oxford: The Clarendon Press). (Reprinted 1981, New York: Dover Publications.)
Hewitt, Edwin and Savage, Leonard J. (1955) 'Symmetric measures on Cartesian products', *Transactions of the American Mathematical Society* **80**, 470–501.
Hume, David (1739) *A Treatise of Human Nature*. London. (Page references are to the 2nd edition of the L. A. Selbe-Bigge text, revised by P. H. Nidditch, Oxford: The Clarendon Press, 1978.)
Hussey, Edward (1972) *The Presocratics* (New York: Charles Scribner's Sons).
Izenman, Alan J. and Zabell, Sandy L. (1981) 'Babies and the blackout: The genesis of a misconception', *Social Science Research* **10**, 282–299.
Jaynes, Edwin T. (1958) *Probability Theory in Science and Engineering*. Colloquium Lectures in Pure and Applied Science, no. 4 (Dallas: Socony Mobil Oil).
Jeffrey, Richard C. (1977) 'Mises redux', *Basic Problems in Methodology and Linguistics: Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, Part III* (R. Butts and J. Hintikka, eds.) (Dordrecht: D. Reidel).
Johnson, William Ernest (1924) *Logic, Part III: The Logical Foundations of Science* (Cambridge University Press).
Johnson, William Ernest (1932) 'Probability: The deductive and inductive problems', *Mind* **49**, 409–423.
Kahn, Charles H. (1960) *Anaximander and the Origins of Greek Cosmology* (New York: Columbia University Press).
Kahneman, D., Slovic, P., and Tversky, A. (1982) *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press).
Keynes, John Maynard (1921) *A Treatise on Probability* (London: Macmillan).
Kirk, G. S. and Raven, J. E. (1957) *The Presocratic Philosophers: A Critical History*

with a Selection of Texts (Cambridge University Press).
Kyburg, Henry (1974) 'Propensities and probabilities', *British Journal for the Philosophy of Science* **25**, 358–375.
Laplace, Pierre Simon Marquis de (1952) *A Philosophical Essay on Probabilities* (F. W. Truscott and F. L. Emory, trans.) (New York: Dover Publications).
Long, A. A. (1974) *Hellenistic Philosophy: Stoics, Epicureans, Sceptics* (New York: Charles Scribner's Sons).
Mill, John Stuart (1843) *A System of Logic*, 2 vols. London.
Michell, J. (1767) 'An inquiry into the probable parallax and magnitude of the fixed stars from the quantity of light which they afford to us, and the particular circumstances of their situation', *Philosophical Transactions of the Royal Society* **57**, 234–264.
Murray, F. H. (1930) 'Note on a scholium of Bayes', *Bulletin of the American Mathematical Society* **36**, 129–132.
Oresme, Nicole (1966) *De proportionnibus proportionum and Ad pauca respicientes* (E. Grant ed. and trans.) (University of Wisconsin Press).
Owen, G. E. L. (1966) 'Plato and Parmenides on the timeless present', *The Monist* **50**, 317–340.
Pearson, Karl (1978) *The History of Statistics in the 17th and 18th Centuries* (E. S. Pearson, ed.) (New York: Macmillan).
Porter, Theodore (1986) *The Rise of Statistical Thinking* (Princeton University Press).
Ramsey, Frank Plumpton (1926) 'Truth and probability', in *The Foundations of Mathematics and Other Logical Essays* (R. B. Braithwaite, ed.) (London: Routledge and Kegan Paul, 1931) pp. 156–198.
Salmon, Wesley C. (1980) 'John Venn's *Logic of Chance*', in *Pisa Conference Proceedings*, vol. 2 (J. Hintikka, D. Gruender, and E. Agazzi, eds.) (Dordrecht: D. Reidel).
Savage, Leonard J. (1954) *The Foundations of Statistics* (New York: John Wiley) (Reprinted 1972, New York: Dover).
Slovic, P., Fischhoff, B., and Lichtenstein, S. (1976) 'Cognitive processes and societal risk taking', in J. S. Carroll and J. W. Payne (eds.), *Cognition and Social Behavior* (Hillsdale, N.J.: Erlbaum).
Stigler, Stephen M. (1982) 'Thomas Bayes's Bayesian inference', *Journal of the Royal Statistical Society* Series A **145**, 250–258.
Stigler, Stephen M. (1986) *The History of Statistics* (Harvard University Press).
Stough, Charlotte L. (1969) *Greek Skepticism* (University of California Press).
Strong, John V. (1976) 'The infinite ballot box of nature: De Morgan, Boole, and Jevons on probability and the logic of induction', *PSA 1976: Proceedings of the Philosophy of Science Association* **1**, 197–211.
Strong, John V. (1978) 'John Stuart Mill, John Herschel, and the "probability of causes"', *PSA 1978: Proceedings of the Philosophy of Science Association*, **1**, 31–41.
Todhunter, Isaac (1865) *A History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace* (London: Macmillan) (Reprinted 1965, New York: Chelsea.)
Venn, John (1858) *The Logic of Chance* (3rd ed.) (London: Macmillan).
Wilson, John Cook (1900) 'Inverse or "a posteriori" probability', *Nature* **63**, 154–156.
von Wright, Georg Henrik (1957) *The Logical Problem of Induction* (2nd revised edition). (New York: Macmillan.)
Zabell, Sandy L. (1982) 'W. E. Johnson's "sufficientness" postulate', *Annals of Statistics* **10**, 1091–1099.

HAIM GAIFMAN

A THEORY OF HIGHER ORDER PROBABILITIES[1]

INTRODUCTION

The assignment of probabilities is the most established way of measuring uncertainties on a quantitative scale. In the framework of subjective probability, the probabilities are interpreted as someone's (the agent's) degrees of belief. Since justified belief amounts to knowledge, the assignment of probabilities, in as much as it can be justified, expresses knowledge. Indeed, knowledge of probabilities appears to be the basic kind of knowledge that is provided by the experimental sciences today.

This is knowledge of a partial, or incomplete, nature, but not in the usual sense of "partial". Usually we mean by "partial knowledge" knowledge of some, but not all, of the facts. But knowing that a given coin is unbiased does not enable one to deduce any non-tautological fact concerning the results of the next, say fifty tosses; every sequence of outcomes is possible. And yet it constitutes very valuable knowledge about these very same outcomes.

What is the objective content of this knowledge? What kind of fact is the fact that the true probability of "heads" is 0.5, i.e., that the coin is unbiased? I have argued elsewhere, (1983), that rather than to classify subjective and objective probabilities as two different kinds we should do better to regard them as two extremes of a spectrum. In that paper I considered the following question: Assuming a probability distribution which represents someone's beliefs, what is it that makes this distribution "objective"? As a way of answering it I pointed out and analyzed two aspects of objectiveness: inner stability and success. To go into these points here would make for too long a digression. So I shall start by taking it for granted that certain probability assignments are regarded by us as expressing fuller knowledge than other assignments. We also think that these "better" assignments are more likely to succeed, or to be in tune with the actual world. I shall not elaborate here on what constitutes "being in tune with the world". Let me only point out that the notion can be given precise meaning, as is illustrated

Brian Skyrms and William L. Harper (eds.), Causation, Chance, and Credence, Vol. I, 191–219.