INTRODUCTION

As will have been clear in Part I, and even more in Part II, probability has become a rich field of study for philosophy. Diverging views about probability also play an ever more important role in philosophical controversy—in general epistemology as well as in philosophy of science proper.

But probability is also a prime area for applications of symmetry arguments. I think these two points are not unconnected. Several times in Part I we came across the idea of a unique probability singled out on purely logical grounds—*logical probability*. In each case I asserted that this does not exist, that it is a philosophical will-o'-the-wisp. Here I return to this point, with a good deal of historical argument to draw on, for it concerns the first use of symmetries in probability theory. It is true that the historical controversy extended into our century, but I regard it as clearly settled now that probability is not uniquely assignable on the basis of a Principle of Indifference, or any other logical grounds.

Similarly in Part II, a crucial role was played by assertions about change of probabilities by Conditionalization. Paradoxically, it appears that this rule has the status of logic, and also that it need not be obeyed. This will be investigated in the last chapter, and central to it is a symmetry argument which fixes the form of admissible rules in probability kinematics.

In the choice of these topics I have been specially concerned to bring us back full circle to earlier parts of this book. The more technical points investigated here—exploiting symmetry arguments—can substantiate or destroy positions taken in general philosophy of science and epistemology. As a result I have ignored much of intrinsic interest—for example, the most famous symmetry result of all, De Finetti's representation theorem for exchangeable (i.e. permutation invariant) probability functions. I have also ignored here to some extent my aim of elucidating the role of symmetry in theory and model construction in physical science. But such omissions can be made good especially fittingly within philosophical discussions of quantum mechanics.

12.
Indifference: The Symmetries of Probability

On estime la probabilité d'un événement par le nombre des cas favourables divisé par le nombre des cas possibles. La difficulté ne consiste que dans l'énumération des cas.

Lagrange, quoted as epigraph to ch. 1 of J. Bertrand, *Calcul des probabilités*.

SINCE its inception in the seventeenth century, probability theory has often been guided by the conviction that symmetry can dictate probability. The conviction is expressed in such slogan formulations as that equipossibility implies equal probability, and honoured by such terms as indifference and sufficient reason. As in science generally we can find here symmetry arguments proper that are truly a priori, as well as arguments that simply assume contingent symmetries, and 'arguments' that reflect the thirst for a hidden, determining reality. The great failure of symmetry thinking was found here, when indifference disintegrated into paradox; and great success as well, sometimes real, sometimes apparent. The story is especially important for philosophy, since it shows the impossibility of the ideal of logical probability.

1. INTUITIVE PROBABILITY

A traveller approaches a river spanned by bridges that connect its shores and islands. There has been a great storm the night before, and each bridge was as likely as not to be washed away. How probable is it that the traveller can still cross? This puzzle, devised by Marcus Moore, clearly depends on the pattern of bridges represented in Figure 12.1.

It also depends on whether the survival of a bridge affects the survival of another. The traveller believes not. Thus for him each bridge had an independent 50 per cent probability of washing away.

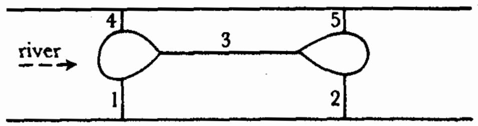


FIG. 12.1. A symmetry argument for probabilities

There is a simple but plodding solution (see *Proofs and illustrations*). But there is a symmetry argument too. Imagine that besides the traveller, there is also a boat moving downstream. The boatman's problem is to get through, which is possible if sufficiently many bridges have been washed away. What is the probability he can get through? Our first observation is that he faces a problem with the same abstract structure. For the traveller, the *entries* are bridges 1 and 2, while for the boatman they are 1 and 4. The *exits* are 4 and 5 for traveller, and are 2 and 5 for boatman. For both there is a *connector*, namely bridge 3. So each sees lying before him the 'maze'

entry exit
connector
entry exit

Good and bad are reversed for traveller and boatman; but suppose that for each, the good state of a bridge has the same independent probability of 50 per cent. Now, by the great Symmetry Requirement, essentially similar problems must have the same solution. Hence:

1. Probability (traveller crosses) = Probability (boat gets through)

But the problems are not only similar; they are also related. For if the traveller has some unbroken path across, the boat cannot get through; and vice versa. Therefore:

2. Probability (boat gets through) = Probability (traveller does not cross)
3. [from 1 and 2] Probability (traveller crosses) = Probability (traveller does not cross)

So it is exactly as likely as not that the traveller will cross—the probability is 50 per cent.

This is a remarkable example, not only as a pure instance of a symmetry argument, but because it introduces all the basic ingredients in the three centuries of controversy over the relation between symmetry and probability. In this problem, the *initial probabilities* are given: 50 per cent for any bridge that it will wash away. We are also given the crucial probability datum about how these eventualities are related: they are *independent*. That means that the collapse of one bridge is neither more nor less probable, on the supposition that some other bridge is washed away. (We are here distinguishing simple probability from *conditional* probability, marked by such terms as 'on the supposition that' or 'given that'.) Then, purely a priori reasoning gives us the probabilities for the events of interest.

The great question for classical probability theory was: can the initial probabilities themselves be deduced too, on the basis of symmetry considerations? If we knew absolutely nothing about storms and bridges, except that one can wash away the other, would rationality not have required us to regard both possible outcomes as equally likely? Once the answer seemed to be obviously *Yes*, and now it seems self-evidently to be *No*, to many of us. But our century also saw the most sophisticated defences of the *yes* answer. And the history of the controversy spun off important and lasting insights.

Proofs and illustrations

In our example, the symmetry transformation used mapped bridge 2 into 4, and vice versa, leaving the others fixed. The *entry-connector-exit* structure is invariant, as is the probability of 'good' (i.e. *whole* for traveller and *broken* for boatman). The reader is invited to consider similar patterns with 1, 3, 4, 5 islands, and to generalize.

The single probability calculus principle that was utilized was—writing '*P*' for 'Probability':

$$P(A) = 1 - P(A)$$

which itself is an immediate corollary to the two axioms

I. $0 = P(\text{contradiction}) \leqslant P(A) \leqslant P(\text{tautology}) = 1$
II. $P(A) + P(B) = P(A \text{ or } B) + P(A \text{ and } B)$

which together exhaust the entire finitary probability theory. For our present purposes, it is not necessary to focus on this calculus (which will be explored further in the next chapter), but the following notions will be relevant (and will be employed intuitively in this chapter):

The *conditional probability* $P(A|B)$ of A *given* that B equals $P(A \text{ and } B)/P(B)$
A and B are (*stochastically* or *statistically*) *independent* exactly if $P(A|B) = P(A)$

That conditional probability $P(A|B)$ is defined only if the *antecedent* B has probability $P(B) \neq 0$. The independence condition is equivalent to

$$P(B|A) = P(B)$$
$$P(A \text{ and } B) = P(A)P(B)$$

always provided the conditional probabilities are defined. The last equation shows clearly, of course, that the condition is symmetric in A and B.

2. CELESTIAL PRIOR PROBABILITIES

The modern history of probability began with the Pascal-Fermat correspondence of 1654. The problems they discussed concerned gambling, games of chance. If someone wanted to draw practical advantage from these studies, he would learn from them how to calculate probabilities of winning (or expectation of gain) from initial probabilities in the gambling set-up. But of course he would have to know those initial probabilities already. While we cannot attribute much sophistication here to the gambler, we may plausibly believe that he takes a hard-nosed empirical stance on this. He believes that the dice are fair exactly if all possible numerical combinations come up equally often—and that this assertion is readily testable even in a small number of tosses. Daggers and rapiers will be drawn if a challenged and tested die comes up even three sixes in a row. We know of course from the play *Rosencrantz and Guildenstern Are Dead* how inconclusive such tests must be on a more sophisticated understanding of probability. But the crucial role and status of initial probability hypotheses appears much more clearly in a different sort of problem.

The Academy of Sciences in Paris proposed a prize subject for 1732 and 1734: the configuration of planetary orbits in our solar system. This configuration may be described as follows: each planet orbits in a plane inclined no more than 7.5° to the sun's equator, and the orbits all have the same direction.[2]

The prize was divided between John Bernoulli and his son Daniel. The latter included three arguments that this configuration cannot be attributed to mere chance. Of these the third argument is a typical eighteenth-century 'calculation' of initial probabilities: 7.5° is $\frac{1}{12}$ of 90° (possible maximum inclination of orbit to equator if we ignore direction); there are six (known) planets, so the probability of this configuration happening 'by chance' is $(\frac{1}{12})^6$, which is negligibly small (*circa* 3 in 10 million).

Daniel Bernoulli has here made two assumptions: of a certain *uniformity* (the probability of at most $\frac{1}{12}$ of the maximum, equals $\frac{1}{12}$) and of *independence* (the joint probability of the six statements is the product of their individual probabilities). Before scrutinizing these assumptions, let us look at two more examples.

Buffon, in his *Historie naturelle* gives an argument similar to Daniel Bernoulli's.[3] Buffon says that the mutual inclination of any two planetary orbits is at most 7.5° Taking direction into account, the maximum is 180°, so the chance of this equals $\frac{1}{24}$. Taking now one planet as fixed, we have five others. The joint probability of all five orbits to be inclined no more than 7.5° is therefore $(\frac{1}{24})^5$. This probability (*circa* 1 in 10 million) is approximately three times smaller than the one noted by Bernoulli. Independently Buffon notes that the probability that all six planets should move in the same west to east direction for us, equals $(\frac{1}{2})^6$. It is clear that he is calculating initial probabilities by the same assumptions as Daniel Bernoulli.

In Laplace's writings on celestial mechanics we find another such example.[4] Bernoulli and Buffon argued for a common origin of the planets, that is, a common cause, on the basis of the improbability of mere chance or coincidence. Laplace argues conversely that a certain fact is not initially improbable, and therefore needs no common-cause explanation. The fact in question was that among the many observed comets, not a single hyperbolic trajectory has been reported.[5] Laplace demonstrates that the probability of a comet with hyperbolic orbit is exceedingly low. The demonstration is based on a uniform distribution of probability over the possible directions of motion of comets entering the sun's gravitational field at some large given distance from the sun.

3. INDIFFERENCE AND SUFFICIENT REASON

It is clear that each of these authors is entertaining what we may call a chance hypothesis: that the phenomenon in question arises 'by mere chance', that is, without the presence of causal or other factors constraining the outcome. There is an ambiguity here: are the probabilities assigned the correct ones (*a*) given *no* hypotheses or assumptions about the physical situation, or (*b*) given a substantial, contingent hypothesis about the absence of certain physical features?

If the former is the case, we have here typical symmetry thinking: the fact that certain information is absent in the statement of the problem, is used as a constraint on the solution. If the latter, we are in the presence of a metaphysical assumption, which may have empirical import: that nature, when certain physical constraints are absent, is equally likely to produce any of the unconstrained possibilities, and therefore tends to produce each equally often.

Ian Hacking locates the first theoretical discussion of this topic in Leibniz's memorandum '*De incerti aestimatione*' (1678).[6] In this note Leibniz equates probability with gradations of possibility ('*probabilitas est gradus possibilitas*'). He states the Principle of Indifference, that equipossible cases have the same probability, and asserts that such a principle can be 'proved by metaphysics'.

We can only speculate what metaphysical proof Leibniz envisaged, but it must surely be based on his Principle of Sufficient Reason. Leibniz's programme set out in the *Discourse of Metaphysics* was to deduce the structure of reality from the nature of God. As a first step, this nature entails that God does, or creates, nothing without sufficient reason. In this marriage of metaphysics with divine epistemology, the difference between points (*a*) and (*b*) above vanishes. For Leibniz's God solves the problem of what nature shall do without contributing factors of his own to destroy the symmetries of the problem-as-stated.

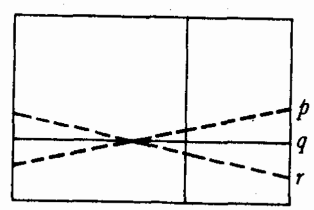
This is how Leibniz must have derived symmetry principles governing nature—determining what the real, objective probabilities shall be in a physical situation. We cannot be sure on the basis of this brief note, but he must have given the principle of sufficient reason also this form: that a rational being should assign equal

probabilities to distinct possibilities unless there be explicit reason to differentiate them. Since Leibniz clearly appreciated the great value of such an equation for metaphysics, he must have appreciated that strictly speaking, his new beginning for metaphysics effects a collapse of two logically distinct problems.

It was certainly in the terminology of sufficient reasons—perhaps always with a equivocation between (*we have reason*) and (*there is reason*)—that principles of indifference were formulated. There were two; we have seen both at work in the arguments of Bernoulli, Buffon, and Laplace.

The first is the *Principle of Uniform Distribution*. Suppose I shoot bullets at a target and am such a poor marksman that it makes no difference at which point of the target I aim. Then any two equal areas on the target are equally likely to be hit. We call this a uniform distribution. The first indifference principle for assigning probabilities is *to assume a uniform distribution in the absence of reasons to the contrary*.

The second is the *Principle of Stochastic Independence*. I explained independence above; let me illustrate it here. Suppose we are told that 40 per cent of the population smokes and 10 per cent has lung cancer. This gives me the probability that a randomly chosen person is a smoker, or has lung cancer, but does not tell me the joint probability of these two characteristics. There are three cases (see Fig. 12.2). Each of the three lines p, q, r has 10 per cent of the area below it. In the case of the horizontal line q, the joint probability of lung cancer *and* smoking is 10 per cent of 40 per cent, namely 4 per cent. For p it is larger and for r it is smaller. The second indifference principle is *to assume statistical independence, in the absence of reasons to the contrary*.



p : positive correlation
q : independence
r : negative correlation

FIG. 12.2

Are these two principles consistent with each other? The joint probability of two events is the same as the ordinary probability of a single complex event. It seems possible therefore that the two principles could be made to apply to the same example, and offer contradictory advice. In the *Proofs and illustrations* we will see that this is not so; the two are consistent with each other.

Proofs and illustrations

Let us consider two variables, say *height* h and *weight* w. Suppose height varies from zero to 10 and weight from zero to 100. Given no other information (hence no reasons to diverge from uniformity or independence), assign probabilities to all possibilities.

The first procedure is to choose uniform distributions for each:

1. $P(0 \leqslant h \leqslant a) = a/10 \qquad P(0 \leqslant w \leqslant b) = b/100$

Then calculate the joint probability by assuming independence:

2. $P(0 \leqslant h \leqslant a \textit{ and } 0 \leqslant w \leqslant b) = (a/10)(b/100)$

The other procedure is to look at the complex variable hw which has pairs of numbers as values. A person with height 6 and weight 60 has hw equal to $<6, 60>$. The big rectangle in Fig. 12.3 encompasses all possibilities ($0 \leqslant h \leqslant 10$ *and* $0 \leqslant w \leqslant 100$) while the smaller one describes the possibility of having hw fall between $<0, 0>$ and $<a, b>$ in the proper sense of 'between'. Uniformity alone applies now and demands a probability proportional to the area:

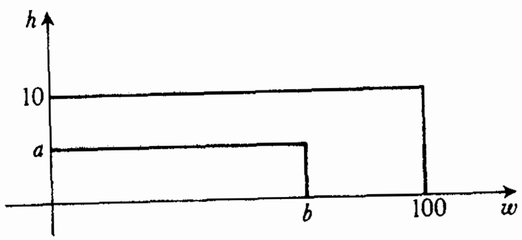


FIG. 12.3

3. $P(<0,0> \leqslant hw \leqslant <a, b>) = ab/1000$

But as we see, 2 and 3 agree. We have proved in effect that if variables h and w are uniformly distributed and independent, then the complex variable hw is uniformly distributed. Hence the two principles are mutually consistent and together constitute the great symmetry principle of classical probability theory—the *Principle of Indifference*.

4. BUFFON'S NEEDLE: EMPIRICAL IMPORT OF INDIFFERENCE

If we must assign initial probabilities, in the absence of relevant information, reason bids us be like Buridan's ass. Do not choose between $P(A) > P(-A)$ and $P(A) < P(-A)$, but set them equal. Similarly in such a case, do not choose between $P(A \text{ and } B) > P(A) \,.\, P(B)$ and $P(A \text{ and } B) < P(A) \,.\, P(B)$, but set those equal as well. Very well; but will nature oblige us with frequencies to which these initial probabilities have a good fit? Is this dictate of reason one that will let reason unlock the mysteries of nature?

An empiricist will ask these questions with a distinct tinge of mockery to his voice. But here we should report a marvellous example in which calculation by the Principle of Indifference led to beautifully confirmed empirical results. This is Buffon's needle problem. It is much more probative than planetary orbit and comet examples, where one only finds *explanation*—that beautiful but airy creature of the fecund imagination—and not *prediction*.

Buffon's needle problem[7]

Given: a large number of parallel lines are drawn on the floor, and a needle is dropped. What is the probability that the needle cuts one of the lines?

To simplify the problem without loss of essential generality, let the lines be exactly two needle lengths apart. Touching will count as cutting, but clearly at most one line is cut. We may even speak sensibly of the line nearest the needle's point (choose either if the point is exactly halfway between). Then our question is equivalent to: what is the probability that the needle cuts this nearest line? In Fig. 12.4 the needle point is a distance $0 \leqslant d \leqslant 1$ away from line L, and its inclination to L is the angle θ. Thus we have:

favourable cases: the needle cuts L exactly if $d \leqslant y = \sin\theta$

This θ varies from zero to 2π ($= 360$ degrees), and so we can diagram the situation with an area of 1 (needle length) by 2π (radians) as in Fig. 12.5. To distinguish the favourable cases from the unfavourable ones, we draw in the sine curve and shade the area where $y \geqslant d$. Assuming independence and uniform distribution, the probability of the favourable cases must be proportional to the

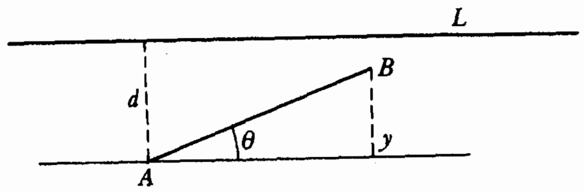


FIG. 12.4. Buffon's needle

shaded area. Since a little calculus quickly demonstrates that this area equals 2, we arrive at the number $2/2\pi$:

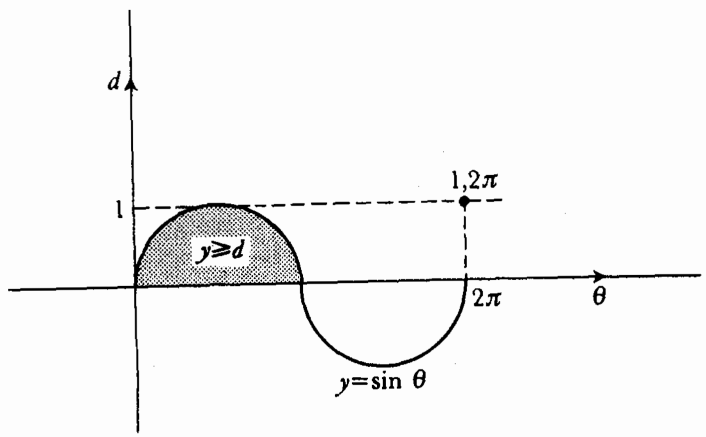


FIG. 12.5. Buffon's probability calculation

The probability of a favourable case equals $1/\pi$, the solution Buffon himself found for his problem.

Since the experiment can be carried out, this is an empirical prediction. It has been carried out a number of times and the outcomes have been in excellent agreement with Buffon's prediction.[8] Now is this not marvellous and a result to make the rationalist metaphysician squeal with delight? For the assumption of symmetry in the probabilities of equipossible cases has here led to a true prediction made a priori.

5. THE CHALLENGE: BERTRAND'S PARADOXES

What I have so far recounted has been very favourable to the Principle of Indifference. Many readers, knowing of its later rejection, but perhaps less familiar with attempts to refine and save it, may already be a little impatient. I will argue for the rejection of its uncritical versions—the empirical phenomena cannot be predicted a priori—but this will be a rejection of naïve symmetry arguments in favour of deeper symmetries, with due respect for the insights that were gained along the way.

We have seen that the Principle has two parts, which are indeed consistent with each other. We have also seen the significant successes of explanations and predictions arrived at in the eighteenth century by means of this Principle. But the challenge to this attempt to calculate initial probabilities on the basis of physical symmetry came exactly from the fundamental principle of symmetry arguments. If two problems are essentially the same, they must receive essentially the same solution. So *a fortiori* if a situation can be equally described in terms of different parameters, we should arrive at the same probabilities if we apply the Principle of Indifference to these other parameters. There will be a logical difficulty—indeed, straightforward inconsistency—if different descriptions of the problem lead via Indifference to distinct solutions.

This logical difficulty with the idea was expounded systematically in a series of paradoxes by Joseph Bertrand at the end of the nineteenth century.[9] Leaving his rather complex geometric examples for *Proofs and Illustrations*, let us turn immediately to a paradigmatic but simple example: the perfect cube factory.[10]

> A precision tool factory produces iron cubes with edge length $\leqslant 2$ cm. What is the probability that a cube has length $\leqslant 1$ cm, given that it was produced by that factory?

A naïve application of the Principle of Indifference consists in choosing length l as parameter and assuming a uniform distribution. The answer is then $\frac{1}{2}$. But the problem could have been stated in different words, but logically equivalent form:

Possible cases	Favourable
edge length $\leqslant 2$	length $\leqslant 1$
area of side $\leqslant 4$	area $\leqslant 1$
volume $\leqslant 8$	volume $\leqslant 1$

Treating each statement of the problem naïvely we arrive at answers $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$. These contradict each other.

The correspondence $l^m \leftrightarrow l^n$, for a parameter l with range $(0, k)$ is one to one, but does not preserve equality of intervals.

Hence uniform distribution on l^m entails non-uniform distribution on l^n. Now sometimes the problem is indeed constrained by symmetries. The cubes example illustrates how these constraints may be so minimal as to leave the set of possible solutions unreduced. More information about the factory could improve the situation. But the Indifference Principle is supposed to fill the gap left by missing information!

Even taken by itself, the example is devastating. But since we shall discuss various attempts to salvage Indifference, it is important to assess two more examples, with somewhat different logical features.

Von Kries posed a problem which is like that of the perfect cube factory, in that several parameters are related by a simple logical transformation. Consider volume and density of a liquid. If mass is set equal to 1, then these parameters are related by:

density = 1/volume; volume = 1/density.

But a uniform distribution on parameter x is automatically non-uniform on $y = (1/x)$. For example,

x is between 1 and 2 exactly if y is between $\frac{1}{2}$ and 1
x is between 2 and 3 exactly if y is between $\frac{1}{3}$ and $\frac{1}{2}$.

Here the two intervals for x are equal in length, but the corresponding ones for y are not. Thus Indifference appears to give us two conflicting probability assignments again.

Von Mises's example of a Bertrand-type paradox concerned a mixture of two liquids, wine and water. We have a glass container, with a mixture of water and wine. To remove division by zero from every inversion, let the following be data:

the glass contains 10 cc of liquid, of which at least 1 cc is water and at least 1 cc is wine.

What is the probability that at least 5 cc is water? Let the parameters be:

a = proportion of wine to total: $(1/10) \leqslant a \leqslant (9/10)$
b = proportion of water to total: $(1/10) \leqslant b \leqslant (9/10)$
x = proportion of wine to water: $(1/9) \leqslant x \leqslant 9$
y = proportion of water to wine: $(1/9) \leqslant y \leqslant 9$

Obviously $b = (1 - a)$, $x = (a/b)$, $y = 1/x$, and $a = x/(1 + x)$, so descriptions of the situation by means of any parameter can be completely translated into any other parameter. It is easy to see that the same problem recurs. Here are two equal intervals for the proportion of wine to total:

a = Proportion of wine to total	x = Proportion of wine to water
4/10	4/6 = 2/3
5/10	5/5 = 1
6/10	6/4 = 3/2

Since $1 - (2/3)$ is not equal to $(3/2) - 1$, it is clear that a uniform distribution on the proportion a entails a non-uniform proportion on proportion x.

In each case the Principle of Uniformity is applied to one perfectly adequate description of the problem. The statements of the problem, both as to sets of possible cases and set of favourable cases, differ only verbally. But the great underlying principle of symmetry thinking is that essentially similar problems must receive the same solution. Thus the attempt to assign uniform distribution on the basis of symmetries in these *statements* of the problem, is drastically misguided—it violates symmetry in a deeper sense.

Most writers commenting on Bertrand have described the problems set by his paradoxical examples as not well posed. In such a case, the problem as initially stated is really not one problem but many. To solve it we must be told *what* is random; which means, *which* events are equiprobable; which means, *which* parameter should be assumed to be uniformly distributed.

But that response asserts that in the absence of further information we have no way to determine the initial probabilities. In other words, this response rejects the Principle of Indifference altogether. After all, if we were told as part of the problem which parameter should receive a uniform distribution, no such Principle would be needed. It was exactly the function of the Principle to turn an incompletely described physical problem into a definite problem in the probability calculus.

There have been different reactions. We have to list Henri Poincaré, E. T. Jaynes, and Rudolph Carnap among the writers who believed that the Principle of Indifference could be refined and sophisticated, and thus saved from paradox.

Proofs and illustrations

The famous chord problem asks for the problem that a stick, tossed randomly on a circle, will mark out a chord of given length. For a definite standard of comparison we inscribe an equilateral triangle ABC in the circle (see Fig. 12.6). However we draw the triangle, it is clear that the separated arcs, like arc AEB, must each be $\frac{1}{3}$ of the circumference. Thus the length of the side of any such triangle is the same. In fact it is $r\sqrt{3}$, where r is the radius, and the point D is exactly halfway along the radius OE.

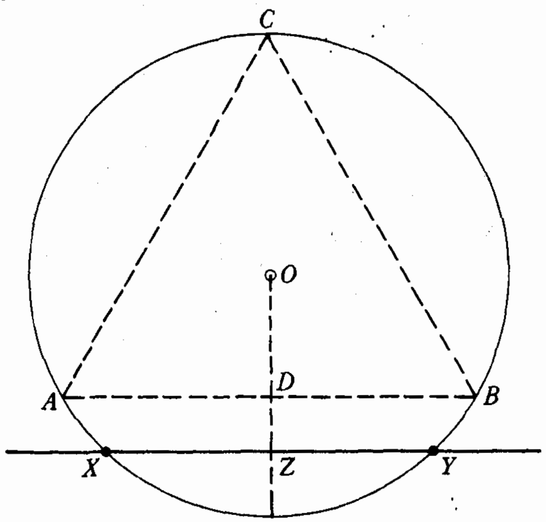


FIG. 12.6. Bertrand's chord problem

What is the probability that chord XY is greater than side AB? If we try to answer this question on the basis of the Principle of Indifference, we actually find three variables which might be asserted to have uniform distribution:

$XY > AB$ exactly if any of the following holds:

(*a*) $OZ < r/2$

(*b*) Y is located between $\frac{1}{3}$ and $\frac{2}{3}$ of the circumference away from X, as measured along the circumference

(*c*) the point Z falls within the 'inner' circle with centre 0 and radius $r/2$.

This gives us three possible applications of the Principle of Uniformity.

Using description (*a*) we reason: OZ can be anything from 0 to r; the interval $[0, r/z]$ of favourable cases has length $\frac{1}{2}$ of the interval $[0, r]$ of possible cases; hence the probability equals $\frac{1}{2}$ (*Solution A*).

Using (*b*) we reason: each point of contact X, Y can be any point on the circle. So given the point X, we can find point Y at any fraction between 0 and 1 of the circumference, measuring counter-clockwise. Of these possible locations, $\frac{1}{3}$ fall in the favourable interval $[\frac{1}{3}, \frac{2}{3}]$; hence the probability equals $\frac{1}{3}$ (*Solution B*).

Using (*c*) we note that the centre Z of the stick can fall anywhere in the whole circle. In the favourable cases it falls in the 'inner' circle with radius $r/2$—which has an area $\frac{1}{4}$ that of the big circle. Hence the probability equals $\frac{1}{4}$ (*Solution C*).

6. SYMMETRIES TO THE RESCUE?

Henri Poincaré and E. T. Jaynes both argued that if we pay attention to the geometric symmetries in Bertrand's problem, we do arrive at a unique solution.[11] Their general idea applies to all apparent ambiguities in the Principle of Indifference: a careful consideration of the exact symmetries of the problem will remove the inconsistency, provided we focus on the symmetry transformations themselves, rather than on the objects transformed.

In order to show the logical structure very clearly I will concentrate on the simple examples of the perfect cubes, mass versus density, and water mixed with the wine. Let us begin by analysing the intuitive reaction to the cube factory, which led us into paradox. Focusing first on the parameter of length, we used the natural length measure for intervals:

$$m(a, b) = b - a$$

This is the *underlying measure*[12] that gave us our probabilities for cases inside the range [0,2]:

$$P(a, b) = m(a,b)/m(\text{range}) \text{ for } a, b \text{ inside the range.}$$

Now this underlying measure has a very special feature, from the point of view of symmetry:

Translation invariance: if $x' = x + k$
then $m(a, b) = m(a', b')$.

Up to multiplication by a scalar, m is the unique measure to have this feature. That is easy to see, because one interval can be moved into another by a translation exactly if they have the same length (and are of the same type: open, half-open, etc.).[13]

The number K represents the scale, if $m' = Km$, because for example the length in inches is numerically 12 times that in feet. It will not affect the probability at all, because it will cancel out (being present in both numerator and denominator in the equation for P in terms of m). We have therefore the following result:

> *Translation invariant measure.* The probability distribution on a real valued parameter x is uniquely determined, if we are given its range and the requirement that it derive from an underlying measure which is translation invariant.

In what sort of example would the given be exactly as required? Suppose I tell you that Peter is a marksman with no skill whatever, and an unknown target. Now I ask you the probability that his bullet will land between 10 and 20 feet from my heart, given that it lands within 20 feet. Treating this formally, I choose a line that falls on both my heart and the impact point of the bullet, coordinatize this line by choosing a point to call zero, and one foot away from it a point to call + 1. I choose a measure m' on this line, call my heart's coordinate X, and calculate

$$m'(X + 10, X + 20)/m'(X, X + 20)$$

and give you the resulting number as answer. If my procedure was properly in tune with the problem, this answer should better not depend on how I chose the points to call zero and + 1 (which two choices together determined the coordinate X). That entails that m' must be translation invariant, and is therefore now uniquely identified. We note with pleasure that the answer is also not affected by the choice of the foot as unit of measurement—as indeed it should not, because nothing in the problem hinged on its Anglo-Saxon peculiarities.

Now, in what sort of problem is the 'given' so different that this procedure is inappropriate? Obviously, when translation invariance is the wrong symmetry. This happens when the range of the physical

quantity in question is not closed under addition and subtraction, for example, if the quantity has an infimum, which acts as natural zero point. For example, no classical object has negative or zero volume, mass, or absolute temperature.

In such a case, the scale or unit may still be irrelevant. For the transformation of the scaling unit consists simply in multiplication by a positive number, which operation does not take us out of this range. Consider now von Kries's problem, which concerns the positive quantities mass and density. With the units of measurement essentially irrelevant we look for an underlying measure

$$M(a, b) = M(ka, kb)$$

for any positive number k (invariance under *dilations*).

There is indeed such a measure, and it is unique in the same sense.[14] That is the *log uniform* distribution:

$$M(a, b) = \log b - \log a$$

where log is the natural logarithm. This function has the nice properties:

$$\log(xy) = \log x + \log y$$
$$\log(x^n) = n\log x$$
$$\log(1) = 0$$

but should be used only for positive quantities, because it moves zero to minus infinity. The first of these equations shows already that M is dilation invariant. The second shows us what is now regarded as equiprobable:

> The intervals (b^n, b^{n+k}) all receive the same value $k\log b$, so if within the appropriate range, the following are series of equiprobable cases:
>
> $(0.1, 1), (1, 10), (10, 100), \ldots, (10^n, 10^{n+1}), \ldots$
> $(0.2, 1), (1, 5), (5, 25), \ldots, (5^n, 5^{n+1}), \ldots$

and so forth. A probability measure derived from the log uniform distribution will therefore always give higher probabilities 'closer' to zero, by our usual reckoning.

For example, in the case of temperature we have since Kelvin accepted that this is essentially a positive quantity. Of course we are at liberty to give the name -273 to absolute zero. But this does not remove the infimum; subtraction eventually takes one outside the range. The presence of this infimum creates, or rather is, an asymmetry: it obstructs translation invariance. But it is no obstacle to dilation invariance, so the log uniform distribution is right—it is dictated by the symmetries of the problem.

This reasoning, being rather abstract, may not get us over our initial feeling of surprise. But as Roger Rosencrantz pointed out, we can test all this on the von Kries problem.[15] Our argument implies that von Kries's puzzle is due to focusing on the wrong transformation group. Attention to the right one dictates use of the log uniform distribution. To our delight this removes the conflict:

$$\begin{aligned} M(1/b \leqslant 1/x \leqslant 1/a) &= K(\log a^{-1} - \log \quad b^{-1}) \\ &= K(\log b - \log a) \\ &= M(a \leqslant x \leqslant b). \end{aligned}$$

This is certainly a success for this approach to Indifference.

Consider next the perfect cube factory. Suppose that again we regard the unit of measurement as essentially irrelevant to this problem, conceived in true generality, but observe that length, area, volume are positive quantities. The uniqueness of the log uniform measure for dilation invariance, forces us then to use it as underlying the correct probabilities. This will not help us, unless we ask all our questions about intervals that exclude *zero*; but for them it works wonderfully well:

> What is the probability that the length is $\leqslant 2$, given that it is between 1 and 3 inclusive?
> What is the probability that the area is $\leqslant 4$, given that it is between 1 and 9 inclusive?
> The probability that the length is $\leqslant 2$, given that it is between 1 and 3 inclusive, equals $M(1, 2)/M(1, 3) = \log 2/\log 3 = 0.631$.
> The probability that the area is $\leqslant 4$, given that it is between 1 and 9 inclusive, equals $M(1, 4)/M(1, 9) = 2\log 2/2\log 3 = 0.631$.

Thus the two equivalent questions do receive the same answer. The point is perfectly general, because the exponent becomes a multiplier, which appears in both numerator and denominator, and so cancels out. This is again a real success. By showing us how to reformulate the problem, and then using its symmetries to determine a unique solution, this approach has as it were taught us how to understand our puzzled but insistent intuitions.

There is therefore good prima-facie reason to take this approach seriously. In the *Proofs and Illustrations* I shall show how this approach does give us a neat solution for the puzzle of Buffon's needle, construed as a Bertrand problem. But in subsequent sections we'll see that the approach does not generalize sufficiently to save the Principle of Indifference.

Proofs and illustrations

I shall here explain this rescue by geometric symmetries with another illustration. For this purpose I choose Buffon's needle problem again, for properly understood it can itself be described as a rudimentary Bertrand paradox.[16].

Buffon assumes no marksmanship—the location of the lines on the floor does not, as far as we know, affect the location of the fallen needle. So our description of the situation utilizes a frame of reference chosen for convenience, in which we treat as X-axis the line through needle point A which is parallel to the drawn lines, as in Fig. 12.4. Here d is the Y-coordinate of line L, θ the inclination of line AB to the X-axis, and y is Y-coordinate of B, and A is the origin.

Why not assume that y and d are independent and uniformly distributed? We must be careful to describe y so that it does not depend on d. But it is just sine θ, and θ does not depend on d, so that is fine. Thus y ranges from -1 to $+1$ (being measured from the X-axis, chosen so that the line L has equation $Y = 1$). The possible and favourable cases are depicted in Fig. 12.7, and we see that the probability of $y \leqslant d$ equals $\frac{1}{4}$. Hence by applying the Principle of Indifference to Buffon's problem *differently but equivalently described*, we have arrived at a different solution.

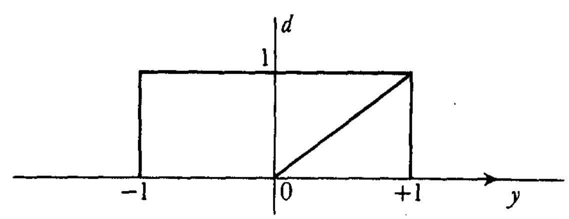


FIG. 12.7

But our description—or rather the solution that utilizes this description in the Principle of Indifference—may itself be faulted for failing to respect geometric symmetry. Consider what happens if the axes are rotated through some angle around point A—that is, the orientation of the lines drawn on the floor is changed. Whatever method of solution we propose, should not make the answer—probability of a cut—depend on this orientation, for the problem remains essentially unchanged. (The aspect varied did not appear at all in the statement of the problem.) How do the two rival solutions vary with respect to this criterion?

Buffon's solution fares very well. For the initial parameter (angle which the needle makes with the X-axis) is changed by adding something (the angle of rotation), modulo 360°. A uniform distribution on that initial parameter induces automatically a uniform distribution also on its transform—equal angular intervals continue to receive equal probability.

But, and here is the rub, if we assume that y is uniformly distributed, it follows that y' (the corresponding coordinate in the rotated frame) is not. The easy way to see this is to look at equal increments in y and notice that they do not correspond to equal increments in y'.

To see this it is necessary to use the formula that transforms coordinates, when the frame is rotated. If the original coordinates of a point are (x, y) they become, upon rotation through angle α around the origin

$$t(x) = x \cos \alpha - y \sin \alpha$$
$$t(y) = x \sin \alpha + y \cos \alpha$$

In our case, point B has coordinates (x, y) but because $AB = 1$ we know that $x^2 + y^2 = 1$. Hence $x = \sqrt{(1 - y^2)}$ and we have

$$t(y) = \sqrt{(1 - y^2)} \sin \alpha + y \cos \alpha$$

Let us now look at two events that have equal probability if y has uniform distribution:

$$[y \, \varepsilon \, (0, 1/2)] \qquad [y \, \varepsilon \, (1/2, 1)]$$

These are the same events as

$$[t(y) \, \varepsilon \, (t(0), t(1/2))] \qquad [t(y) \, \varepsilon \, (t(1/2), t(1))]$$

If the variable $t(y)$ has uniform distribution, these events will have equal probabilities only if

$$t(1/2) - t(0) = t(1) - t(1/2)$$

so that is what we need to check. A single counter-example will do, so let us choose the angle of 30° (i.e. $\pi/6$ radians) which has sine $\frac{1}{2}$ and cosine $(\sqrt{3})/2$. Therefore:

$$t(0) = \sin(\tfrac{1}{6}\pi) = \tfrac{1}{2}$$
$$t(\tfrac{1}{2}) = (\sqrt{3})/2 \sin \tfrac{1}{6}\pi + \tfrac{1}{2} \cos \tfrac{1}{6}\pi$$
$$= (\sqrt{3})/2\,(\tfrac{1}{2}) + \tfrac{1}{2}((\sqrt{3})/2) = (\sqrt{3})/2$$
$$t(1) = \cos \tfrac{1}{6}\pi = (\sqrt{3})/2$$

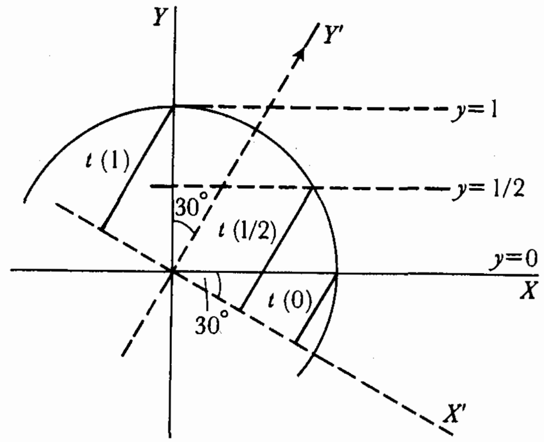


FIG. 12.8

It is very obvious that our desired equation does not hold. Figure 12.8 shows the different cases.

7. PYRRHIC VICTORY AND ULTIMATE DEFEAT

The successes we found in the preceding section, even together with their more sophisticated variants (to be discussed in the *Proofs and Illustrations*), constitute only a Pyrrhic victory. Again we can see this in simple examples, just because of the power of the uniqueness results utilized. Recall that invariance under translations and invariance under dilations each dictate an essentially *unique* answer to all probability questions. What happens when the examples take on more structure?

Peter Milne, writing about Rosencrantz's solution to von Kries's problem, has shown exactly how things go wrong.[17] To show this, he asked how the above results are to be applied to von Mises's water and wine problem. Let us again ask the same question in two different ways, referring back to the notation we had before.

What is the probability that at least 5 cc is water?
b = proportion of water to the total
x = proportion of wine to water.
$P(b \geqslant 0.5 \mid 0.1 \leqslant b \leqslant 0.9 = (\log 0.9 - \log 0.5)/(\log 0.9 - \log 0.1) = 0.267$
$P(x \leqslant 1 \mid 1/9 < x \leqslant 9) = \log 1 - \log(1/9))/(\log 9 - \log(1/9)) = \log 9/2\log 9 = 0.5$

We have received two contradictory answers.

Were we justified in treating the problem in this way? Well, the problem specified cc as unit of measurement but we have just as much warrant to regard this as irrelevant as we had for cm in the cubes problem. If we focus on parameter x here, say, we must treat it in the same way, if we have indeed found the correct form of the Principle of Indifference. Restating the problem then in terms of b, we have not introduced any new information—so we must derive the answer from the probability distribution on x, plus the definition of b in terms of x. Exactly the same would apply if we had started with b, and then moved on to x. But the two end results are not the same, so we have our paradox back.

It is also rather easy to see the pattern that will produce such paradoxes. A translation invariant measure will be well behaved with respect to addition and multiplication, while a dilation invariant measure will be equally good with respect to multiplication and exponentiation. But the relation between b and x uses both sorts of operations:

b = water/(water + wine) = water/10
x = wine/water = (10 − water)/water
hence, water = $10b$ and $x = (100 - b)/b$.

Now neither sort of measure will do. If the required dilation invariance did not dictate an essentially unique measure, we would perhaps have had some leeway to look for something other than logarithms—but we do not.

The history of the Principle of Indifference is instructive. If its mention in scientific sermons serves to remind us to look for symmetries, then it serves well. But a rule to determine initial probabilities a priori it is not. It violates a higher symmetry requirement when it is conceived of in that way.

Even if the Principle were unambiguous, the question whether its results would be probability functions with a good fit to actual frequencies in nature, would anyway be a purely contingent one. To imagine that it would not be—that empirical predictions could be made a priori, by 'pure thought' analysis—is feasible only on the assumption of some metaphysical scheme such as Leibniz's, in which the symmetries of the problems which God selects for attention, determine the structure of reality.[18] But because it is not unambiguous, even that assumption would leave us stranded, unless we knew how God selected his problems.

When E. T. Jaynes[19] discussed Bertrand's chord paradox, although noting that most writers had regarded it as an ill-posed problem, he responded:

> But do we really believe that it is beyond our power to predict by 'pure thought' the results of such a simple experiment? The point at issue is far more important than merely resolving a geometric puzzle; for . . . applications of probability theory to physical experiments usually lead to problems of just this type. . . . (p. 478)

Jaynes's analysis of the Bertrand chord problem is along the lines of the preceding section. He shows that there is only one solution which derives from a measure which is invariant under Euclidean transformations.

But when we look more carefully at other parts of Jaynes's paper we see that his more general conclusions nullify the radical tone. Jaynes says of von Mises's water and wine problem, that the fatal ambiguities of the Principle of Indifference remain. More important: the strongest conclusion Jaynes manages to reach is merely one of *advice*, to regard a problem as having a definite solution until the contrary has been proved. The method he advises us to follow is that of symmetry arguments:

> To summarize the above results: if we merely specify complete ignorance, we cannot hope to obtain any definite prior distribution, because such a statement is too vague to define any mathematically well-defined problem. We are defining what we mean by complete ignorance far more precisely if we can specify a set of operations which we recognize as transforming

the problem into an equivalent one, and the desideratum of consistency then places non-trivial restrictions on the form of the prior.[20]

But as we know, this method always rests on assumptions which may or may not fit the physical situation in reality. Hence it cannot lead to a priori predictions. Success, when achieved, must be attributed to the good fortune that nature fits and continues to fit the general model with which the solution begins.

Proofs and illustrations

Harold Jeffreys introduced the search for invariant priors into the foundations of statistics; there has been much subsequent work along these lines by others.[21] We must conclude with Dawid, however, that the programme 'produces a whole range of choices in some problems, and no prior free from all objections in others' ('Invariant Prior Distributions', 235). I just wish to take up here the elegant logical analysis that Jaynes introduced to generalize the approach which we have been studying in these last two sections.[22]

Here some powerful mathematical theorems come to our aid. For under certain conditions, there exists indeed only one possible probability assignment to a group, so there is no ambiguity.

The general pattern of the approach I have been outlining is as follows. First one selects the correct group of transformations of our sets K which should leave the probability measure invariant. Call the group G. Then one finds the correct probability measure p on this group. Next define

(1) $P(A) = p(\{g \text{ in } G\colon g(x_0)\ \varepsilon A\})$

where x_0 is a chosen reference point in the set K on which we want our probability defined. If everything has gone well, P is the probability measure 'demanded' by the group.

What is required at the very least is that (*a*) p is a privileged measure on the group; (*b*) P is invariant under the action of the group; and (*c*) P is independent of the choice of x_0. Mathematics allows these desiderata to be satisfied: if the group G has some 'nice' properties, and we require p to be a left Haar measure (which means that $p(S) = p(\{gg' : g'\ \varepsilon\ S\})$ for any part S of G and any member g) then these desired consequences follow, and p, P are essentially unique.[23]

This is a very tight situation, and the required niceties can be

expected in geometric models such as are used to define Bertrand's chord problem. But other sorts of models will not be equally nice; and even if they are, different models of the same situation could fairly bring us diverse answers. In any case there is a no a priori reason why all phenomena should fit models with such 'nice' properties only.

8. THE ETHICS OF AMBIGUITY

From the initial example, of a traveller on a treacherous shore, to the partial but impressive successes in the search for invariant priors, I have tried to emphasize how much symmetry considerations tell us. That is the positive side of this definitive dissolution of the idea of unique logical probability. Yet the story is far from complete, and its tactical and strategic suggestions for model construction far from exhausted.

But throughout the history of this subject, there wafts the siren melody of empirical probabilities determined a priori on the basis of pure symmetry considerations. The correct appreciation leads us to exactly the same conclusion as in Chapter 10. Once a problem is modelled, the symmetry requirement may give it a unique, or at least greatly constrained solution. The modelling, however, involves substantive assumptions: an implicit selection of certain parameters as alone relevant, and a tacit assumption of structure in the parameter space. Whenever the consequent limitations are ignored, paradoxes bring us back to our senses—symmetries respected in one modelling of the problem *entail* symmetries broken in another model. As soon as we took the first step, symmetries swept us along in a powerful current—but nature might have demanded a different first step, or embarkation in a different stream.

Facts are ambiguous. It is vain to desire prescience: which resolution of the present ambiguities will later facts vindicate? Our models of the facts, on the other hand, are not ambiguous; they had better not be. To choose one, is therefore a risk. To eliminate the risk is to cease theorizing altogether. That is one message of these paradoxes.