CHAPTER 10

PHYSICS OF "RANDOM EXPERIMENTS"

"I believe, for instance, that it would be very difficult to persuade an intelligent physicist that current statistical practice was sensible, but that there would be much less difficulty with an approach via likelihood and Bayes' theorem." — G. E. P. Box (1962).

As we have noted several times, the idea that probabilities are physically real things, based ultimately on observed frequencies of random variables, underlies most recent expositions of probability theory, which would seem to make it a branch of experimental science. At the end of Chapter 8 we saw some of the difficulties that this view leads us to; in some real physical experiments the distinction between random and nonrandom quantities is so obscure and artificial that you have to resort to black magic in order to force this distinction into the problem at all. But that discussion did not reach into the serious physics of the situation. In this Chapter, we take time off for an interlude of physical considerations that show the fundamental difficulty with the notion of "random" experiments.

An Interesting Correlation

There have always been dissenters from the "frequentist" view who have maintained, with Laplace, that probability theory is properly regarded as the "calculus of inductive reasoning," and is not fundamentally related to random experiments at all. A major purpose of the present work is to demonstrate that probability theory can deal, consistently and usefully, with far more than frequencies in random experiments, if only it is allowed to do so. According to this view, consideration of random experiments is only one specialized *application* of probability theory, and not even the most important one; for probability theory as logic solves far more general problems of reasoning which have nothing to do with chance or randomness, but a great deal to do with the real world. In the present Chapter we carry this further and show that 'frequentist' probability theory has major logical difficulties in dealing with the very random experiments for which it was invented.

One who studies the literature of these matters perceives that there is a strong correlation; those who have advocated the non–frequency view have tended to be physicists, while up until very recently mathematicians, statisticians, and philosophers almost invariably favored the frequentist view. Thus it appears that the issue is not merely one of philosophy or mathematics; in some way not yet clear, it also involves physics.

The mathematician tends to think of a random experiment as an abstraction – really nothing more than a sequence of numbers. To define the "nature" of the random experiment he introduces statements – variously termed assumptions, postulates, or axioms – which specify the sample space and assert the existence, and certain other properties, of limiting frequencies.

But in the real world, a random experiment is not an abstraction whose properties can be defined at will. It is surely subject to the laws of physics; yet recognition of this is conspicuously missing from frequentist expositions of probability theory. Even the phrase 'laws of physics' is not to be found in them. But defining a probability as a frequency is not merely an excuse for ignoring the laws of physics; it is more serious than that. We want to show that maintenance of a frequency interpretation to the exclusion of all others *requires* one to ignore virtually all the professional knowledge that scientists have about real phenomena. If the aim is to draw inferences about real phenomena, this is hardly the way to begin.

As soon as a specific random experiment is described, it is the nature of a physicist to start thinking, not about the abstract sample space thus defined, but about the physical mechanism of the phenomenon being observed. The question whether the usual postulates of probability theory are compatible with the known laws of physics is capable of logical analysis, with results that have a direct bearing on the question, not of the mathematical consistency of frequency and non–frequency theories of probability, but of their applicability in real situations. In our opening quotation, the statistician G. E. P. Box noted this; let us analyze his statement in the light both of history and of physics.

Historical Background

As we know, probability theory started in consideration of gambling devices by Gerolamo Cardano in the 16'th Century, and by Pascal and Fermat in the 17'th; but its development beyond that level, in the 18'th and 19'th centuries, was stimulated by applications in astronomy and physics, and was the work of people – James and Daniel Bernoulli, Laplace, Poisson, Legendre, Gauss, Boltzmann, Maxwell, Gibbs – most of whom we would describe today as mathematical physicists.

But reactions against Laplace started already in the mid Nineteenth Century, when Cournot, Ellis, Boole, and Venn – none of whom had any training in physics – were unable to comprehend Laplace's rationale and attacked what he did, simply ignoring all his successful results. In particular, John Venn, a philosopher without the tiniest fraction of Laplace's knowledge of either physics or mathematics, nevertheless considered himself competent to write scathing, sarcastic attacks on Laplace's work. In Chapter 16 we note his possible later influence on the young R. A. Fisher.

Boole (1854, Chapters XX and XXI) shows repeatedly that he does not understand the function of Laplace's prior probabilities (to represent a *state of knowledge* rather than a physical fact). In other words, he too suffers from the Mind Projection Fallacy. On p. 380 he rejects a uniform prior probability assignment as 'arbitrary' and explicitly *refuses* to examine its consequences; by which tactics he prevents himself from learning what Laplace was really doing and why.

Laplace was defended staunchly by the mathematician Augustus de Morgan and the physicist W. Stanley Jevons,† who understood Laplace's motivations and for whom his beautiful mathematics was a delight rather than a pain. Nevertheless, the attacks of Boole and Venn found a sympathetic hearing in England among non–physicists. Perhaps this was because biologists, whose training in physics and mathematics was for the most part not much better than Venn's, were trying to find empirical evidence for Darwin's theory and realized that it would be necessary to collect and analyze large masses of data in order to detect the small, slow trends that they visualized as the means by which evolution proceeds. Finding Laplace's mathematical works too much to digest, and since the profession of Statistician did not yet exist, they would naturally welcome suggestions that they need not read Laplace after all.

In any event, a radical change took place at about the beginning of this Century when a new group of workers, not physicists, entered the field. They were concerned mostly with biological problems and with Venn's encouragement proceeded to reject virtually everything done by Laplace. To fill the vacuum, they sought to develop the field anew based on entirely different principles in which one assigned probabilities only to data and to nothing else. Indeed, this did simplify the mathematics at first, because many of the problems solvable by Laplace's methods now lay outside the ambit of their methods. As long as they considered only relatively simple problems (technically, problems with sufficient statistics but no nuisance parameters), the shortcoming was

† Jevons did so many things that it is difficult to classify him by occupation. Zabell (1989), apparently guided by the title of one of his books (1874), describes Jevons as a logician and philosopher of science; from examination of his other works we are inclined to list him rather as a physicist who wrote extensively on economics.

not troublesome. This extremely aggressive school soon dominated the field so completely that its methods have come to be known as "orthodox" statistics, and the modern profession of Statistician has evolved mostly out of this movement.

Simultaneously with this development, the physicists – with Sir Harold Jeffreys as almost the sole exception – quietly retired from the field, and statistical analysis disappeared from the physics curriculum. This disappearance has been so complete that, if today someone were to take a poll of physicists, we think that not one in a hundred could identify such names as Fisher, Neyman, Wald; or such terms as maximum likelihood, confidence interval, analysis of variance.

This course of events – the leading role of physicists in development of the original Bayesian methods, and their later withdrawal from orthodox statistics – was no accident. As further evidence that there is some kind of basic conflict between orthodox statistical doctrine and physics, we may note that two of the most eloquent proponents of non–frequency definitions in the early 20'th Century – Poincaré and Jeffreys – were mathematical physicists of the very highest competence, as was Laplace. Professor Box's statement thus has a clear basis in historical fact.

But what is the nature of this conflict? What is there in the physicist's knowledge that leads him to reject the very thing that the others regard as conferring "objectivity" on probability theory? To see where the difficulty lies, we examine a few simple random experiments from the physicist's viewpoint. The facts we want to point out are so elementary that one cannot believe they are really unknown to modern writers on probability theory. The continual appearance of new textbooks which ignore them merely illustrates what we physics teachers have always known; you can teach a student the laws of physics, but you cannot teach him the art of recognizing the *relevance* of this knowledge, much less the habit of actually *applying* it, in his everyday problems.

How to Cheat at Coin and Die Tossing

Cramér (1946) takes it as an axiom that "Any random variable has a unique probability distribution." From the later context, it is clear that what he really means is that it has a unique *frequency* distribution. If one assumes that the number obtained by tossing a die is a random variable, this leads to the conclusion that the frequency with which a certain face comes up is a physical property of the die; just as much so as its mass, moment of inertia, or chemical composition. Thus, Cramér (p. 154) states:

> "The numbers p_r should, in fact, be regarded as physical constants of the particular die that we are using, and the question as to their numerical values cannot be answered by the axioms of probability theory, any more than the size and the weight of the die are determined by the geometrical and mechanical axioms. However, experience shows that in a well–made die the frequency of any event r in a long series of throws usually approaches 1/6, and accordingly we shall often assume that all the p_r are equal to 1/6 $\cdots$."

To a physicist, this statement seems to show utter contempt for the known laws of mechanics. The results of tossing a die many times do *not* tell us any definite number characteristic only of the die. They tell us also something about how the die was tossed. If you toss "loaded" dice in different ways, you can easily alter the relative frequencies of the faces. With only slightly more difficulty, you can still do this if your dice are perfectly "honest."

Although the principles will be just the same, it will be simpler to discuss a random experiment with only two possible outcomes per trial. Consider, therefore, a "biased" coin, about which I. J. Good (1962) has remarked:

> "Most of us probably think about a biased coin as if it had a physical probability. Now whether it is defined in terms of frequency or just falls out of another type of theory, I think we do argue that way. I suspect that even the most extreme subjectivist such as de Finetti would have to agree that he did sometimes think that way, though he would perhaps avoid doing it in print."

We do not know de Finetti's private thoughts, but would observe that it is just the famous exchangeability theorem of de Finetti which shows us how to carry out a probability analysis of the biased coin *without* thinking in the manner suggested.

In any event, it is easy to show how a physicist would analyze the problem. Let us suppose that the center of gravity of this coin lies on its axis, but displaced a distance x from its geometrical center. If we agree that the result of tossing this coin is a "random variable," then according to the axiom stated by Cramér and hinted at by Good, there must exist a definite functional relationship between the frequency of heads and x:

$$p_H = f(x)\ . \qquad (10\text{–}1)$$

But this assertion goes far beyond the mathematician's traditional range of freedom to invent arbitrary axioms, and encroaches on the domain of physics; for the laws of mechanics are quite competent to tell us whether such a functional relationship does or does not exist.

The easiest game to analyze turns out to be just the one most often played to decide such practical matters as the starting side in a football game. Your opponent first calls "heads" or "tails" at will. You then toss the coin into the air, catch it in your hand, and without looking at it, show it first to your opponent, who wins if he has called correctly. It is further agreed that a "fair" toss is one in which the coin rises at least nine feet into the air, and thus spends at least 1.5 seconds in free flight.

The laws of mechanics now tell us the following. The ellipsoid of inertia of a thin disc is an oblate spheroid of eccentricity $1/\sqrt{2}$. The displacement x does not affect the symmetry of this ellipsoid, and so according to the Poinsot construction, as found in textbooks on rigid dynamics [such as Routh (1955) or Goldstein (1980, Chapter 5)], the polhodes remain circles concentric with the axis of the coin. In consequence, the character of the tumbling motion of the coin while in flight is exactly the same for a biased as an unbiased coin, except that for the biased one it is the center of gravity, rather than the geometrical center, which describes the parabolic "free particle" trajectory.

An important feature of this tumbling motion is conservation of angular momentum; during its flight the angular momentum of the coin maintains a fixed direction in space (but the angular *velocity* does not; and so the tumbling may appear chaotic to the eye). Let us denote this fixed direction by the unit vector n; it can be any direction you choose, and it is determined by the particular kind of twist you give the coin at the instant of launching. Whether the coin is biased or not, it will show the same face throughout the motion if viewed from this direction (unless, of course, n is exactly perpendicular to the axis of the coin, in which case it shows no face at all).

Therefore, in order to know which face will be uppermost in your hand, you have only to carry out the following procedure. Denote by k a unit vector passing through the coin along its axis, with its point on the "heads" side. Now toss the coin with a twist so that k and n make an acute angle, then catch it with your palm held flat, in a plane normal to n. On successive tosses, you can let the direction of n, the magnitude of the angular momentum, and the angle between n and k, vary widely; the tumbling motion will then appear entirely different to the eye on different tosses, and it would require almost superhuman powers of observation to discover your strategy.

Thus, anyone familiar with the law of conservation of angular momentum can, after some practice, cheat at the usual coin–toss game and call his shots with 100 per cent accuracy. You can obtain any frequency of heads you want; and *the bias of the coin has no influence at all on the results*!

Of course, as soon as this secret is out, someone will object that the experiment analyzed is too "simple." In other words, those who have postulated a physical probability for the biased coin have, without stating so, really had in mind a more complicated experiment in which some kind of "randomness" has more opportunity to make itself felt.

While accepting this criticism, we cannot suppress the obvious comment: scanning the literature of probability theory, isn't it curious that so many mathematicians, usually far more careful than physicists to list all the qualifications needed to make a statement correct, should have failed to see the need for any qualifications here? However, to be more constructive, we can just as well analyze a more complicated experiment.

Suppose that now, instead of catching the coin in our hands, we toss it onto a table, and let it spin and bounce in various ways until it comes to rest. Is this experiment sufficiently "random" so that the true "physical probability" will manifest itself? No doubt, the answer will be that it is not sufficiently random if the coin is merely tossed up six inches starting at the table level, but it will become a "fair" experiment if we toss it up higher.

Exactly how high, then, must we toss it before the true physical probability can be measured? This is not an easy question to answer, and we make no attempt to answer it here. It would appear, however, that anyone who asserts the existence of a physical probability for the coin ought to be prepared to answer it; otherwise it is hard to see what content the assertion has (that is, there is no way to confirm it or disprove it).

We do not deny that the bias of the coin will now have some influence on the frequency of heads; we claim only that the amount of that influence depends very much on how you toss the coin so that, again in this experiment, there is no definite number $p_H = f(x)$ describing a physical property of the coin. Indeed, even the direction of this influence can be reversed by different methods of tossing, as follows.

However high we toss the coin, we still have the law of conservation of angular momentum; and so we can toss it by *Method A*: to ensure that heads will be uppermost when the coin first strikes the table, we have only to hold it heads up, and toss it so that the total angular momentum is directed vertically. Again, we can vary the magnitude of the angular momentum, and the angle between n and k, so that the motion appears quite different to the eye on different tosses, and it would require very close observation to notice that heads remains uppermost throughout the free flight. Although what happens after the coin strikes the table is complicated, the fact that heads is uppermost at first has a strong influence on the result, which is more pronounced for large angular momentum.

Many people have developed the knack of tossing a coin by *Method B*: it goes through a phase of standing on edge and spinning rapidly about a vertical axis, before finally falling to one side or the other. If you toss the coin this way, the eccentric position of the center of gravity will have a dominating influence, and render it practically certain that it will fall always showing the same face. Ordinarily, one would suppose that the coin prefers to fall in the position which gives it the lowest center of gravity; *i.e.,* if the center of gravity is displaced toward tails, then the coin should have a tendency to show heads. However, for an interesting mechanical reason, which we leave for you to work out from the principles of rigid dynamics, method B produces the opposite influence, the coin strongly preferring to fall so that its center of gravity is high.

On the other hand, the bias of the coin has a rather small influence in the opposite direction if we toss it by *Method C*: the coin rotates about a horizontal axis which is perpendicular to the axis of the coin, and so bounces until it can no longer turn over.

In this experiment also, therefore, a person familiar with the laws of mechanics can toss a biased coin so that it will produce predominantly either heads or tails, at will. Furthermore, the effect of method A persists whether the coin is biased or not; and so one can even do this with a perfectly "honest" coin. Finally, although we have been considering only coins, essentially the same mechanical considerations (with more complicated details) apply to the tossing of any other object, such as a die.

The writer has never thought of a biased coin 'as if it had a physical probability' because, being a professional physicist, I know that it does *not* have a physical probability. From the fact that we have seen a strong preponderance of heads, we cannot conclude legitimately that the coin is biased; it may be biased, or it may have been tossed in a way that systematically favors heads. Likewise, from the fact that we have seen equal numbers of heads and tails, we cannot conclude legitimately that the coin is "honest." It may be honest, or it may have been tossed in a way that nullifies the effect of its bias.

Experimental Evidence

Since the conclusions just stated are in direct contradiction to what is postulated, almost universally, in expositions of probability theory, it is worth noting that you can verify them easily in a few minutes of experimentation in your kitchen. An excellent "biased coin" is provided by the metal lid of a small pickle jar, of the type which is not knurled on the outside, and has the edge rolled inward rather than outward, so that the outside surface is accurately round and smooth, and so symmetrical that on an edge view one cannot tell which is the top side.

Suspecting that many people not trained in physics, simply would not believe the things just claimed without experimental proof, we have performed these experiments with a jar lid of diameter $d = 2\ 5/8$ inches, height $h = 3/8$ inch. Assuming a uniform thickness for the metal, the center of gravity should be displaced from the geometrical center by a distance $x = dh/(2d + 8h) = 0.120$ inches; and this was confirmed by hanging the lid by its edge and measuring the angle at which it comes to rest. Ordinarily, one expects this bias to make the lid prefer to fall bottom side (*i.e.,* the inside) up; and so this side will be called "heads." The lid was tossed up about 6 feet, and fell onto a smooth linoleum floor. I allowed myself ten practice tosses by each of the three methods described, and then recorded the results of a number of tosses by: method A deliberately favoring heads, method A deliberately favoring tails, method B, and method C, as given in Table 10.1.

Method	*No. of tosses*	*No. of heads*
A(H)	100	99
A(T)	50	0
B	100	0
C	100	54

Table 10.1. Results of tossing a "biased coin" in four different ways.

In method A the mode of tossing completely dominated the result (the effect of bias would, presumably, have been greater if the "coin" were tossed onto a surface with a greater coefficient of friction). In method B, the bias completely dominated the result (in about thirty of these tosses it looked for a while as if the result were going to be heads, as one might naively expect; but each time the "coin" eventually righted itself and turned over, as predicted by the laws of rigid dynamics). In method C, there was no significant evidence for any effect of bias. The conclusions are pretty clear.

A holdout can always claim that tossing the coin in any of the four specific ways described is "cheating," and that there exists a "fair" way of tossing it, such that the "true" physical probabilities of the coin will emerge from the experiment. But again, the person who asserts this should be prepared to define precisely what this fair method is, otherwise the assertion is without content. Presumably, a fair method of tossing ought to be some kind of random mixture of methods A(H), A(T), B, C, and others; but what is a "fair" relative weighting to give them? It is difficult to see how one could define a "fair" method of tossing except by the condition that it should result in a certain frequency of heads; and so we are involved in a circular argument.

This analysis can be carried much further, as we shall do below; but perhaps it is sufficiently clear already that analysis of coin and die tossing is not a problem of abstract statistics, in which one is free to introduce postulates about "physical probabilities" which ignore the laws of physics. It is a problem of mechanics, highly complicated and irrelevant to probability theory except insofar as it forces us to think a little more carefully about how probability theory must be formulated if it is to be applicable to real situations. Performing a random experiment with a coin does not tell us what the physical probability of heads is; it may tell us something about the bias, but it also tells us something about how the coin is being tossed. Indeed, unless we know how it is being tossed, we cannot draw any reliable inferences about its bias from the experiment.

It may not, however, be clear from the above that conclusions of this type hold quite generally for random experiments, and in no way depend on the particular mechanical properties of coins and dice. In order to illustrate this, consider an entirely different kind of random experiment, as a physicist views it.

Bridge Hands

Elsewhere we quote Professor Wm. Feller's pronouncements on the use of Bayes' theorem in quality control testing (Chap.17), on Laplace's rule of succession (Chap. 19), and on Daniel Bernoulli's conception of the utility function for decision theory (Chap. 13). He does not fail us here either; in this interesting textbook (Feller, 1951), he writes: "*The number of possible distributions of cards in bridge is almost* 10^{30}*. Usually, we agree to consider them as equally probable. For a check of this convention more than* 10^{30} *experiments would be required—a billion of billion of years if every living person played one game every second, day and night.*" Here again, we have the view that bridge hands possess "physical probabilities," that the uniform probability assignment is a "convention," and that the ultimate criterion for its correctness must be observed frequencies in a random experiment.

The thing which is wrong here is that none of us – not even Feller – would be willing to use this criterion with a real deck of cards. Because, if we know that the deck is an honest one, our common sense tells us something which carries more weight than 10^{30} random experiments do. We would, in fact, be willing to accept the result of the random experiment *only if it agreed with our preconceived notion that all distributions are equally likely.*

To many, this last statement will seem like pure blasphemy – it stands in violent contradiction to what we have all been taught is the correct attitude toward probability theory. Yet in order to see why it is true, we have only to imagine that those 10^{30} experiments *had* been performed, and the uniform distribution was not forthcoming. If all distributions of cards have equal frequencies, then any combination of two specified cards will appear together in a given hand, on the average, once in $(52 \times 51)/(13 \times 12) = 17$ deals. But suppose that the combination (Jack of hearts – Seven of clubs) appeared together in each hand three times as often as this. Would we then accept it as an established fact that there is something about the particular combination (Jack of hearts – Seven of clubs) that makes it inherently more likely than others?

We would not. We would reject the experiment and say that the cards had not been properly shuffled. But once again we are involved in a circular argument, because there is no way to define a "proper" method of shuffling except by the condition that it should produce all distributions with equal frequency!

But any attempt to find such a definition involves one in even deeper logical difficulties; one dare not describe the procedure of shuffling in exact detail because that would destroy the "randomness" and make the exact outcome predictable and always the same. In order to keep the experiment "random", one must describe the procedure incompletely, so that the outcome will be different on different runs. But how could one prove that an incompletely defined procedure will produce all distributions with equal frequency? It seems to us that the attempt to uphold Feller's postulate of physical probabilities for bridge hands leads one into an outright logical contradiction.

Conventional teaching holds that probability assignments must be based fundamentally on frequencies; and that any other basis is at best suspect, at worst irrational with disastrous consequences. On the contrary, this example shows very clearly that *there is a principle for determining probability assignments which has nothing to do with frequencies, yet is so compelling that it takes precedence over any amount of frequency data.* If present teaching does not admit the existence of this principle, it is only because our intuition has run so far ahead of logical analysis – just as it does in elementary geometry – that we have never taken the trouble to present that logical analysis in a mathematically respectable form. But if we learn how to do this, we may expect to find that the mathematical formulation can be applied to a much wider class of problems, where our intuition alone would hardly suffice.

In carrying out a probability analysis of bridge hands, are we really concerned with physical probabilities; or with inductive reasoning? To help answer this, consider the following scenario. The date is 1956, when the writer met Willy Feller and had a discussion with him about these matters. Suppose I had told him that I have dealt at bridge 1000 times, shuffling "fairly" each time; and that in every case the seven of clubs was in my own hand. What would his reaction be? He would, I think, mentally visualize the number

$$\left(\frac{1}{4}\right)^{1000} = 10^{-602} \tag{10–2}$$

and conclude instantly that I have not told the truth; and no amount of persuasion on my part would shake that judgment. But what accounts for the strength of his belief? Obviously, it cannot be justified if our assignment of equal probabilities to all distributions of cards (therefore probability 1/4 for the seven of clubs to be in the dealer's hand) is merely a "convention," subject to change in the light of experimental evidence; he rejects my reported experimental evidence, just as we did above. Even more obviously, he is not making use of any knowledge about the outcome of an experiment involving 10^{30} bridge hands.

Then *what is the extra evidence he has*, which his common sense tells him carries more weight than do any number of random experiments; but whose help he refuses to acknowledge in writing textbooks? In order to maintain the claim that probability theory is an experimental science, based fundamentally not on logical inference but on frequency in a random experiment, it is necessary to suppress some of the information which is available. This suppressed information, however, is just what enables our inferences to approach the certainty of deductive reasoning in this example and many others.

The suppressed evidence is, of course, simply our recognition of the *symmetry* of the situation. The only difference between a seven and an eight is that there is a different number printed on the face of the card. Our common sense tells us that where a card goes in shuffling depends only on the mechanical forces that are applied to it; and not on which number is printed on its face. If we observe any systematic tendency for one card to appear in the dealer's hand, which persists on indefinite repetitions of the experiment, we can conclude from this only that there is some systematic tendency in the procedure of shuffling, which alone determines the outcome of the experiment.

Once again, therefore, performing the experiment tells you nothing about the "physical probabilities" of different bridge hands. It tells you something about how the cards are being shuffled. But the full power of symmetry as cogent evidence has not yet been revealed in this argument; we return to it presently.

General Random Experiments

In the face of all the foregoing arguments, one can still take the following position (as a member of the audience did after one of the writer's lectures): "You have shown only that coins, dice, and cards represent exceptional cases, where physical considerations obviate the usual probability postulates; *i.e.,* they are not really 'random experiments.' But that is of no importance because these devices are used only for illustrative purposes; in the more dignified random experiments which merit the serious attention of the scientist, there *is* a physical probability."

To answer this we note two points. First, we reiterate that when anyone asserts the existence of a physical probability in any experiment, then the onus is on him to define the exact circumstances in which this physical probability can be measured; otherwise the assertion is without content.

This point needs to be stressed: those who assert the existence of physical probabilities do so in the belief that this establishes for their position an 'objectivity' that those who speak only of a 'state of knowledge' lack. Yet to assert as fact something which cannot be either proved or disproved by observation of facts, is the opposite of objectivity; it is to assert something that one could not possibly know to be true. Such an assertion is not even entitled to be called a description of a 'state of knowledge'.

Secondly, note that any specific experiment for which the existence of a physical probability is asserted, is subject to physical analysis like the ones just given, which will lead eventually to an understanding of its mechanism. But as soon as this understanding is reached, then this new experiment will also appear as an exceptional case like the above ones, where physical considerations obviate the usual postulates of physical probabilities.

For, as soon as we have understood the mechanism of any experiment E, then there is logically no room for any postulate that various outcomes possess physical probabilities; for the question: "What are the probabilities of various outcomes $(O_1, O_2 \cdots)$?" then reduces immediately to the question: "What are the probabilities of the corresponding initial conditions $(I_1, I_2 \cdots)$ that lead to these outcomes?"

We might suppose that the possible initial conditions $\{I_k\}$ of experiment E themselves possess physical probabilities. But then we are considering an antecedent random experiment E', which produces conditions I_k as its possible outcomes: $I_k = O'_k$. We can analyze the physical mechanism of E' and as soon as this is understood, the question will revert to: "What are the probabilities of the various initial conditions I'_k for experiment E'?"

Evidently, we are involved in an infinite regress $\{E,\ E',\ E'', \cdots\}$; the attempt to introduce a physical probability will be frustrated at every level where our knowledge of physical law permits us to analyze the mechanism involved. The notion of "physical probability" must retreat continually from one level to the next, as knowledge advances.

We are, therefore, in a situation very much like the "warfare between science and theology" of earlier times. For several centuries, theologians with no factual knowledge of astronomy, physics, biology, and geology, nevertheless considered themselves competent to make dogmatic factual assertions which encroached on the domains of those fields – which they were later forced to retract one by one in the face of advancing knowledge.

Clearly, probability theory ought to be formulated in a way that avoids factual assertions properly belonging to other fields, and which will later need to be retracted (as is now the case for many assertions in the literature concerning coins, dice, and cards). It appears to us that the only formulation which accomplishes this, and at the same time has the analytical power to deal with the current problems of science, is the one which was seen and expounded on intuitive grounds by Laplace and Jeffreys. Its validity is a question of logic, and does not depend on any physical assumptions.

As we saw in Chapter 2, a major contribution to that logic was made by R. T. Cox (1946), (1961), who showed that those intuitive grounds can be replaced by theorems. We think it is no accident that Richard Cox was also a physicist (Professor of Physics and Dean of the Graduate School at Johns Hopkins University), to whom the things we have pointed out here would be evident from the start.

The Laplace–Jeffreys–Cox formulation of probability theory does not require us to take one reluctant step after another down that infinite regress; it recognizes that anything which – like the child's spook – continually recedes from the light of detailed inspection, can exist only in our imagination. Those who believe most strongly in physical probabilities, like those who believe in astrology, never seem to ask what would constitute a controlled experiment capable of confirming or disproving their belief.

Indeed, the examples of coins and cards should persuade us that such controlled experiments are in principle impossible. Performing any of the so–called random experiments will not tell us what the "physical probabilities" are, because *there is no such thing as a "physical probability".* The experiment tells us, in a very crude and incomplete way, something about how the initial conditions are varying from one repetition to another.

A much more efficient way of obtaining this information would be to observe the initial conditions directly. However, in many cases this is beyond our present abilities; as in determining the safety and effectiveness of a new medicine. Here the only fully satisfactory approach would be to analyze the detailed sequence of chemical reactions that follow the taking of this medicine, in persons of every conceivable state of health. Having this analysis one could then predict, for each individual patient, exactly what the effect of the medicine will be.

Such an analysis being entirely out of the question at present, the only feasible way of obtaining information about the effectiveness of a medicine is to perform a "random" experiment. No two patients are in exactly the same state of health; and the unknown variations in this factor constitute the variable initial conditions of the experiment, while the sample space comprises the set of distinguishable reactions to the medicine. Our use of probability theory in this case is a standard example of inductive reasoning which amounts to the following:

If the initial conditions of the experiment (*i.e.,* the physiological conditions of the patients who come to us) continue in the future to vary over the same unknown range as they have in the past, then the relative frequency of cures will, in the future, approximate those which we have observed in the past. In the absence of positive evidence giving a reason why there should be some change in the future, *and* indicating in which direction this change should go, we have no grounds for predicting any change in either direction, and so can only suppose that things will continue in more or less the same way. As we observe the relative frequencies of cures and side–effects to remain stable over longer and longer times, we become more and more confident about this conclusion. But this is only inductive reasoning – there is no deductive proof that frequencies in the future will not be entirely different from those in the past.

Suppose now that the eating habits or some other aspect of the life style of the population starts to change. Then the state of health of the incoming patients will vary over a different range than before, and the frequency of cures for the same treatment may start to drift up or down. Conceivably, monitoring this frequency could be a useful indicator that the habits of the population are changing, and this in turn could lead to new policies in medical procedures and public health education.

At this point, we see that the logic invoked here is virtually identical with that of industrial quality control, discussed in Chapter 4. But looking at it in this greater generality makes us see the role of induction in science in a very different way than has been imagined by some philosophers.

Induction Revisited

As we noted in Chapter 9, some philosophers have rejected induction on the grounds that there is no way to prove that it is "right" (theories can never attain a high probability); but this misses the point. The function of induction is to tell us, not which predictions are right, but which predictions are indicated by our present knowledge. If the predictions succeed, then we are pleased and become more confident of our present knowledge; but we have not learned much.

The real role of induction in science was pointed out clearly by Harold Jeffreys (1931, Chapter 1) over sixty years ago; yet to the best of our knowledge no mathematician or philosopher has ever taken the slightest note of what he had to say:

> "A common argument for induction is that induction has always worked in the past and therefore may be expected to hold in the future. It has been objected that this is itself an inductive argument and cannot be used in support of induction. What is hardly ever mentioned is that induction has often failed in the past and that progress in science is very largely the consequence of direct attention to instances where the inductive method has led to incorrect predictions."

Put more strongly, it is only when our inductive inferences are wrong that we learn new things about the real world. For a scientist, therefore, the quickest path to discovery is to examine those situations where it appears most likely that induction from our present knowledge will fail. But those inferences must be our *best* inferences, which make full use of all the knowledge we have. One can always make inductive inferences that are wrong in a useless way, merely by ignoring cogent information.

Indeed, that is just what Popper did. His trying to interpret probability itself as expressing physical causation not only cripples the applications of probability theory in the way we saw in Chapter 3 (it would prevent us from getting about half of all conditional probabilities right because they express logical connections rather than causal physical ones) – it leads one to conjure up imaginary causes while ignoring what was already known about the real physical causes at work. This can reduce our inferences to the level of pre–scientific, uneducated superstition even when we have good data.

Why do physicists see this more readily than others? Because, having created this knowledge of physical law, we have a vested interest in it and want to see it preserved and used. Frequency or propensity interpretations start by throwing away practically all the professional knowledge that we have labored for Centuries to get. Those who have not comprehended this are in no position to discourse to us on the philosophy of science or the proper methods of inference.

But What About Quantum Theory?

Those who cling to a belief in the existence of "physical probabilities" may react to the above arguments by pointing to quantum theory, in which physical probabilities appear to express the most fundamental laws of physics. Therefore let us explain why this is another case of circular reasoning. We need to understand that present quantum theory uses entirely different standards of logic than does the rest of science.

In biology or medicine, if we note that an effect E (for example, muscle contraction, phototropism, digestion of protein) does not occur unless a condition C (nerve impulse, light, pepsin) is present, it seems natural to infer that C is a necessary causative agent for E. Most of what is known in all fields of science has resulted from following up this kind of reasoning. But suppose that condition C does not always lead to effect E; what further inferences should a scientist draw? At this point the reasoning formats of biology and quantum theory diverge sharply.

In the biological sciences one takes it for granted that in addition to C there must be some other causative factor F, not yet identified. One searches for it, tracking down the assumed cause by a process of elimination of possibilities that is sometimes extremely tedious. But persistence pays off; over and over again medically important and intellectually impressive success has been achieved, the conjectured unknown causative factor being finally identified as a definite chemical compound. Most enzymes, vitamins, viruses, and other biologically active substances owe their discovery to this reasoning process.

In quantum theory, one does not reason in this way. Consider, for example, the photoelectric effect (we shine light on a metal surface and find that electrons are ejected from it). The experimental fact is that the electrons do not appear unless light is present. So light must be a causative factor. But light does not always produce ejected electrons; even though the light from a unimode laser is present with absolutely steady amplitude, the electrons appear only at particular times that are not determined by any known parameters of the light. Why then do we not draw the obvious inference, that in addition to the light there must be a second causative factor, still unidentified, and the physicist's job is to search for it?

What is done in quantum theory today is just the opposite; when no cause is apparent one simply postulates that no cause exists – ergo, the laws of physics are indeterministic and can be expressed only in probability form. The central dogma is that the light determines, not whether a photoelectron will appear, but only the probability that it will appear. The mathematical formalism of present quantum theory – incomplete in the same way that our present knowledge is incomplete – does not even provide the vocabulary in which one could ask a question about the real cause of an event.

Biologists have a mechanistic picture of the world because, being trained to believe in causes, they continue to search for them and find them. Quantum physicists have only probability laws because for two generations we have been indoctrinated not to believe in causes – and so we have stopped looking for them. Indeed, any attempt to search for the causes of microphenomena is met with scorn and a charge of professional incompetence and 'obsolete mechanistic materialism'. Therefore, to explain the indeterminacy in current quantum theory we need not suppose there is any indeterminacy in Nature; the mental attitude of quantum physicists is already sufficient to guarantee it.

This point also needs to be stressed, because most people who have not studied quantum theory on the full technical level are incredulous when told that it does not concern itself with causes; and indeed, it does not even recognize the notion of 'physical reality.' The currently taught interpretation of the mathematics is due to Niels Bohr, who directed the Institute for Theoretical Physics in Copenhagen; therefore it has come to be called 'The Copenhagen Interpretation'.

As Bohr stressed repeatedly in his writings and lectures, present quantum theory can answer only questions of the form: "If this experiment is performed, what are the possible results and their probabilities?" It cannot, as a matter of principle, answer any question of the form: "What is really happening when $\cdots$?" Again, the mathematical formalism of present quantum theory, like Orwellian *newspeak*, does not even provide the vocabulary in which one could ask such a question. These points have been explained in some detail in recent articles (Jaynes, 1986d, 1989, 1990a, 1991c).

We suggest, then, that those who try to justify the concept of 'physical probability' by pointing to quantum theory, are entrapped in circular reasoning, not basically different from that noted above with coins and bridge hands. Probabilities in present quantum theory express the incompleteness of human knowledge just as truly as did those in classical statistical mechanics; only its origin is different.

In classical statistical mechanics, probability distributions represented our ignorance of the true microscopic coordinates – ignorance that was avoidable in principle but unavoidable in practice, but which did not prevent us from predicting reproducible phenomena, just because those phenomena are independent of the microscopic details.

In current quantum theory, probabilities express the ignorance due to our failure to search for the real causes of physical phenomena. This may be unavoidable in practice, but in our present state of knowledge we do not know whether it is unavoidable in principle; the "central dogma" simply asserts this, and draws the conclusion that belief in causes, and searching for them, is philosophically naïve. If everybody accepted this and abided by it, no further advances in understanding of physical law would ever be made; indeed, no such advance has been made since the 1927 Solvay Congress in which this mentality became solidified into physics.‡

But to many people, these ideas are almost impossible to comprehend because they are so radically different from what we have all been taught from childhood. Therefore let us show how just the same situation could have happened in coin tossing, had classical physicists used the same standards of logic that are now used in quantum theory.

Mechanics Under the Clouds

We are fortunate that the principles of Newtonian mechanics could be developed and verified to great accuracy by studying astronomical phenomena, where friction and turbulence do not complicate what we see. But suppose the Earth were, like Venus, enclosed perpetually in thick clouds. The very existence of an external universe would be unknown for a long time, and to develop the laws of mechanics we would be dependent on the observations we can make locally.

Since tossing of small objects is nearly the first activity of every child, it would be observed very early that they do not always fall with the same side up, and that all one's efforts to control the outcome are in vain. The natural hypothesis would be that it is the volition of the object tossed, not the volition of the tosser, that determines the outcome; indeed, that is the hypothesis that small children make when questioned about this.

Then it would be a major discovery, once coins had been fabricated, that they tend to show both sides about equally often; and the equality appears to get better as the number of tosses increases. The equality of heads and tails would be seen as a fundamental law of physics; symmetric objects have a symmetric volition in falling (as indeed, Cramér and Feller seem to have thought).

With this beginning, we could develop the mathematical theory of object tossing, discovering the binomial distribution, the absence of time correlations, the limit theorems, the combinatorial frequency laws for tossing of several coins at once, the extension to more complicated symmetric objects like dice, etc. All the experimental confirmations of the theory would consist of more and more tossing experiments, measuring the frequencies in more and more elaborate scenarios. From such experiments, nothing would ever be found that called into question the existence of that volition of the object tossed; they only enable one to confirm that volition and measure it more and more accurately.

Then suppose that someone was so foolish as to suggest that the motion of a tossed object is determined, not by its own volition, but by laws like those of Newtonian mechanics, governed by its initial position and velocity. He would be met with scorn and derision; for in all the existing experiments there is not the slightest evidence for any such influence. The Establishment would proclaim that, since all the observable facts are accounted for by the volition theory, it is philosophically naïve and a sign of professional incompetence to assume or search for anything deeper. In this respect, the elementary physics textbooks would read just like our present quantum theory textbooks.

Indeed, anyone trying to test the mechanical theory would have no success; however carefully he tossed the coin (not knowing what we know) it would persist in showing head and tails about

‡ Of course, physicists continued discovering new particles and calculation techniques – just as an astronomer can discover a new planet and a new algorithm to calculate its orbit, without any advance in his basic understanding of celestial mechanics.

equally often. To find any evidence for a causal instead of a statistical theory, would require control over the initial conditions of launching, orders of magnitude more precise than anyone can achieve by hand tossing. We would continue almost indefinitely, satisfied with laws of physical probability and denying the existence of causes for individual tosses external to the object tossed – just as quantum theory does today – because those probability laws account correctly for everything that we can observe reproducibly with the technology we are using.

But after thousands of years of triumph of the statistical theory, someone finally makes a machine which tosses coins in absolutely still air, with very precise control of the exact initial conditions. Magically, the coin starts giving unequal numbers of heads and tails; the frequency of heads is being controlled partially by the machine. With development of more and more precise machines, one finally reaches a degree of control where the outcome of the toss can be predicted with 100% accuracy. Belief in "physical probabilities" expressing a volition of the coin is recognized finally as an unfounded superstition. The existence of an underlying mechanical theory is proved beyond question; and the long success of the previous statistical theory is seen as due only to the lack of control over the initial conditions of the tossing.

Because of recent spectacular advances in the technology of experimentation, with increasingly detailed control over the initial states of individual atoms [see, for example, Rempe, *et al* (1987); Knight (1987)], we think that the stage is going to be set, before very many more years have passed, for the same thing to happen in quantum theory; a Century from now the true causes of microphenomena will be known to every schoolboy and, to paraphrase Seneca, they will be incredulous that such clear truths could have escaped us throughout the 20'th Century.

More On Coins and Symmetry

Now we go into a more careful, detailed discussion of some of these points, alluding to technical matters that must be explained more fully elsewhere. The rest of this Chapter is not for the casual reader; only the one who wants a deeper understanding than is conveyed by the above simple scenarios. But many of the attacks on Laplace arise from failure to comprehend the following points.

The problems in which intuition compels us most strongly to a uniform probability assignment are not the ones in which we merely apply a principle of "equal distribution of ignorance." Thus, to explain the assignment of equal probabilities to heads and tails on the grounds that we "saw no reason why either face should be more likely than the other," fails utterly to do justice to the reasoning involved. The point is that we have not merely "equal ignorance." We also have *positive knowledge of the symmetry* of the problem; and introspection will show that when this positive knowledge is lacking, so also is our intuitive compulsion toward a uniform distribution. In order to find a respectable mathematical formulation we therefore need to find first a more respectable verbal formulation. We suggest that the following verbalization does do justice to the reasoning, and shows us how to generalize the principle.

"I perceive here two different problems, Having formulated one definite problem – call it P_1 – involving the coin, the operation which interchanges heads and tails transforms the problem into a different one – call it P_2. If I have positive knowledge of the symmetry of the coin, then I know that all relevant dynamical or statistical considerations, however complicated, are exactly the same in the two problems. Whatever state of knowledge I had in P_1, I must therefore have exactly the same state of knowledge in P_2, except for the interchange of heads and tails. Thus, whatever probability I assign to heads in P_1, consistency demands that I assign the *same* probability to tails in P_2.

Now it might be quite reasonable to assign probability 2/3 to heads, 1/3 to tails in P_1; whereupon from symmetry it must be 2/3 to tails, 1/3 to heads in P_2. This might be the case, for example, if P_1 specified that the coin is to be held between the fingers heads up, and dropped just one inch onto a table. Thus symmetry of the coin by no means compels us to assign equal probabilities to heads and tails; the question necessarily involves the other conditions of the problem.

But now suppose the statement of the problem is changed in just one respect; we are no longer told whether the coin is held initially with heads up or tails up. In this case, our intuition suddenly takes over with a compelling force, and tells us that we *must* assign equal probabilities to heads and tails; and in fact, we must do this *regardless of what frequencies have been observed in previous repetitions of the experiment.*

The great power of symmetry arguments lies just in the fact that they are not deterred by any amount of complication in the details. The conservation laws of physics arise in this way; thus conservation of angular momentum for an arbitrarily complicated system of particles is a simple consequence of the fact that the Lagrangian is invariant under space rotations. In current theoretical physics, almost the only known exact results in atomic and nuclear structure are those which we can deduce by symmetry arguments, using the methods of group theory.

These methods could be of the highest importance in probability theory also, if orthodox ideology did not forbid their use. For example, they enable us, in many cases, to extend the principle of indifference to find consistent prior probability assignments in a continuous parameter space Θ, where its use has always been considered ambiguous. The basic point is that a consistent principle for assigning prior probabilities must have the property that it assigns equivalent priors to represent equivalent states of knowledge.

The prior distribution must therefore be invariant under the symmetry group of the problem; and so the prior can be specified arbitrarily only in the so–called "fundamental domain" of the group (Wigner, 1959). This is a subspace $\Theta_0 \subset \Theta$ such that (1) applying two different group elements $g_i \neq g_j$ to Θ_0, the subspaces $\Theta_i \equiv g_i\,\Theta_0$, $\Theta_j \equiv g_j\,\Theta_0$ are disjoint; and (2) carrying out all group operations on Θ_0 just generates the full hypothesis space: $\cup_j\,\Theta_j = \Theta$.

For example, let points in a plane be defined by their polar coordinates (r, α). If the group is the four-element one generated by a 90° rotation of the plane, then any sector 90° wide, such as $(\beta \leq \alpha < \beta + \pi/2)$ is a fundamental domain. Specifying the prior in any such sector, symmetry under the group then determines the prior everywhere in the plane.

If the group contains a continuous symmetry operation, the dimensionality of the fundamental domain is less than that of the parameter space; and so the probability density need be specified only on a set of points of measure zero, whereupon it is determined everywhere. If the number of continuous symmetry operations is equal to the dimensionality of the space Θ, the fundamental domain reduces to a single point, and the prior probability distribution is then uniquely determined by symmetry alone, just as it is in the case of an honest coin. Later we shall formalize and generalize these symmetry arguments.

There is still an important constructive point to be made about the power of symmetry arguments in probability theory. To see it, let us go back for a closer look at the coin–tossing problem. The laws of mechanics determine the motion of the coin, as describing a certain trajectory in a twelve–dimensional phase space [three coordinates (q_1, q_2, q_3) of its center of mass, three Eulerian angles (q_4, q_5, q_6) specifying its orientation, and six associated momenta $(p_1, \ldots, p_6)$]. The difficulty of predicting the outcome of a toss arises from the fact that very small changes in the location of the initial phase point can change the final results.

Imagine the possible initial phase points to be labelled H or T, according to the final results. Contiguous points labelled H comprise a set which is presumably twisted about in the twelve–dimensional phase space in a very complicated, convoluted way, parallel to and separated by similar T–sets.

Consider now a region R of phase space, which represents the accuracy with which a human hand can control the initial phase point. Because of limited skill, we can be sure only that the initial point is somewhere in R, which has a phase volume

$$\Gamma(R) = \int_R dq_1 \cdots dq_6 \, dp_1 \cdots dp_6$$

If the region R contains both H and T domains, we cannot predict the result of the toss. But what probability should we assign to heads? If we assign equal probability to equal phase volumes in R, this is evidently the fraction $p_H \equiv \Gamma(H)/\Gamma(R)$ of phase volume of R that is occupied by H domains. This phase volume Γ is the "invariant measure" of phase space. The cogency of invariant measures for probability theory will be explained later; for now we note that the measure Γ is invariant under a large group of "canonical" coordinate transformations, and also under the time development, according to the equations of motion. This is Liouville's theorem, fundamental to statistical mechanics; the exposition of Gibbs (1902) devotes the first three Chapters to discussion of it, before introducing probabilities.

Now if we have positive knowledge that the coin is perfectly "honest," then it is clear that the fraction $\Gamma(H)/\Gamma(R)$is very nearly 1/2, and becomes more accurately so as the size of the individual H and T domains become smaller compared to R. Because, for example, if we are launching the coin in a region R where the coin makes fifty complete revolutions while falling, then a one percent change in the initial angular velocity will just interchange heads and tails by the time the coin reaches the floor. Other things being equal, (all dynamical properties of the coin involve heads and tails in the same manner), this should just reverse the final result.

A change in the initial "orbital" velocity of the coin, which results in a one percent change in the time of flight, should also do this (strictly speaking, these conclusions are only approximate, but we expect them to be highly accurate, and to become more so if the changes become less than one percent). Thus, if all other initial phase coordinates remain fixed, and we vary only the initial angular velocity $\dot{\theta}$ and upward velocity $\dot{z}$, the H and T domains will spread into thin ribbons, like the stripes on a zebra. From symmetry, the width of adjacent ribbons must be very nearly equal.

This same "parallel ribbon" shape of the H and T domains presumably holds also in the full phase space.† This is quite reminiscent of Gibbs' illustration of fine–grained and coarse–grained probability densities, in terms of the stirring of colored ink in water. On a sufficiently fine scale, every phase region is either H or T; the probability of heads is either zero or unity. But on the scale of sizes of the "macroscopic" region R corresponding to ordinary skills, the probability density is the coarse–grained one, which from symmetry must be very nearly 1/2 if we know that the coin is honest.

What if we don't consider all equal phase volumes within R as equally likely? Well, it doesn't really matter if the H and T domains are sufficiently small. "Almost any" probability density which is a smooth, continuous function within R, will give nearly equal weight to the H and T domains, and we will still have very nearly 1/2 for the probability of heads. This is an example of a general phenomenon, discussed by Poincaré, that in cases where small changes in initial conditions produce big changes in the final results, our final probability assignments will be, for all practical purposes, independent of the initial ones.

As soon as we know that the coin has perfect dynamical symmetry between heads and tails – *i.e.*, its Lagrangian function

† Actually, if the coin is tossed onto a perfectly flat and homogeneous level floor and is not only perfectly symmetrical under the reflection operation that interchanges heads and tails, but also perfectly round, the probability of heads is independent of five of the twelve coordinates, so we have this intricate structure only in a seven–dimensional space. Let the reader for whom this is a startling statement think about it hard, to see why symmetry makes five coordinates irrelevant (they are the two horizontal coordinates of its center of mass, the direction of its horizontal component of momentum, the Eulerian angle for rotation about a vertical axis, and the Eulerian angle for rotation about the axis of the coin).

$$L(q_1 \ldots p_6) = (\text{Kinetic energy}) - (\text{Potential energy})$$

is invariant under the symmetry operation that interchanges heads and tails – then we know an exact result. No matter where in phase space the initial region R is located, for every H domain there is a T domain of equal size and identical shape, in which heads and tails are interchanged. Then if R is large enough to include both, we shall persist in assigning probability 1/2 to heads.

But now suppose the coin is biased. The above argument is lost to us, and we expect that the phase volumes of H and T domains within R are no longer equal. In this case, the "frequentist" tells us that there still exists a definite "objective" frequency of heads, $p_H \neq 1/2$ which is a measurable physical property of the coin. Let us understand clearly what this implies. *To assert that the frequency of heads is a physical property only of the coin, is equivalent to asserting that the ratio $v(H)/v(R)$ is independent of the location of region R.* If this were true, it would be an utterly unprecedented new theorem of mechanics, with important implications for physics which extend far beyond coin tossing.

Of course, no such thing is true. From the three specific methods of tossing the coin discussed above which correspond to widely different locations of the region R, it is clear that the frequency of heads will depend very much on how the coin is tossed. Method A uses a region of phase space where the individual H and T domains are large compared to R, so human skill is able to control the result. Method B uses a region where, for a biased coin, the T domain is very much larger than either R or the H domain. Only method C uses a region where the H and T domains are small compared to R, making the result unpredictable from knowledge of R.

It would be interesting to know how to calculate the ratio $v(H)/v(R)$ as a function of the location of R from the laws of mechanics; but it appears to be a very difficult problem. Note, for example, that the coin cannot come to rest until its initial potential and kinetic energy have been either transferred to some other object or dissipated into heat by frictional forces; so all the details of how that happens must be taken into account. Of course, it would be quite feasible to do controlled experiments which measure this ratio in various regions of phase space. But it seems that the only person who would have any use for this information is a professional gambler.

Clearly, our reason for assigning probability 1/2 to heads when the coin is honest is not based merely on observed frequencies. How many of us can cite a single experiment in which the frequency 1/2 was established under conditions we would accept as significant? Yet none of us hesitates a second in choosing the number 1/2. Our real reason is simply common–sense recognition of the *symmetry* of the situation. *Prior information which does not consist of frequencies is of decisive importance in determining probability assignments even in this simplest of all random experiments.*

Those who adhere publicly to a strict frequency interpretation of probability jump to such conclusions privately just as quickly and automatically as anyone else; but in so doing they have violated their basic premise that (probability) $\equiv$ (frequency); and so in trying to justify this choice they must suppress any mention of symmetry, and fall back on remarks about assumed frequencies in random experiments which have, in fact, never been performed.†

Here is an example of what one loses by so doing. From the result of tossing a die, we cannot tell whether it is symmetrical or not. But if we know, from direct physical measurements, that the die *is* perfectly symmetrical and we accept the laws of mechanics as correct, then it is no longer plausible inference, but deductive reasoning, that tells us this: *any nonuniformity in the frequencies of different faces is proof of a corresponding nonuniformity in the method of tossing.* The qualitative

† Or rather, whenever anyone has tried to perform such experiments under sufficiently controlled conditions to be significant, the expected equality of frequencies is *not* observed. The famous experiments of Weldon and Wolf are discussed elsewhere in this work.

nature of the conclusions we can draw from the random experiment depend on whether we do or do not know that the die is symmetrical.

This reasoning power of arguments based on symmetry has led to great advances in physics for sixty years; as noted, it is not very exaggerated to say that the only known exact results in mathematical physics are the ones that can be deduced by the methods of group theory from symmetry considerations. Although this power is obvious once noted and it is used intuitively by every worker in probability theory, it has not been widely recognized as a legitimate formal tool in probability theory.‡

We have just seen that in the simplest of the random experiments, any attempt to define a probability merely as a frequency involves us in the most obvious logical difficulties as soon as we analyze the mechanism of the experiment. In many situations where we can recognize an element of symmetry our intuition readily takes over and suggests an answer; and of course it is the same answer that our basic desideratum – that equivalent states of knowledge should be represented by equivalent probability assignments – requires for consistency.

But in situations in which we have positive knowledge of symmetry are rather special ones among all those faced by the scientist. How can we carry out consistent inductive reasoning in situations where we do not perceive any clear element of symmetry? This is an open–ended problem because there is no end to the variety of different special circumstances that might arise. As we shall see, the principle of Maximum Entropy gives a useful and versatile tool for many such problems. But in order to give a start toward understanding this, let's go way back to the beginning and consider the tossing of the coin still another time, in a different way.

Independence of Tosses

"When I toss a coin the probability of heads is one half." What do we mean by this statement? Over the past two centuries millions of words have been written about this simple question. A recent exchange (Edwards, 1991) shows that it is still enveloped in total confusion in the minds of some. But by and large, the issue is between the following two interpretations:

A: "The available information gives me no reason to expect heads rather than tails, or vice versa – I am completely unable to predict which it will be."

B: "If I toss the coin a very large number of times, in the long run heads will occur about half the time – in other words, the frequency of heads will approach 1/2."

We belabor still another time, what we have already stressed many times before: Statement (A) does not describe any property of the coin, but only thr robot's *state of knowledge* (or if you prefer, of ignorance). (B) is, at least by implication, asserting something about the coin. Thus (B) is a very much stronger statement than (A). Note, however, that (A) does not in any way contradict (B); on the contrary, (A) could be a consequence of (B). For if our robot were told that this coin has in the past given heads and tails with equal frequency, this would give it no help at all in predicting the result of the next toss.

Why, then, has interpretation (A) been almost universally rejected by writers on probability and statistics for two generations? There are, we think, two reasons for this. In the first place, there is a widespread belief that if probability theory is to be of any use in applications, we must be able to interpret our calculations in the strong sense of (B). But this is simply untrue, as we have demonstrated throughout the last eight Chapters. We have seen examples of almost all

‡ Indeed, L. J. Savage (1962, p. 102) rejects symmetry arguments, thereby putting his system of 'personalistic' probability in the position of recognizing the need for prior probabilities but refusing to admit any formal principles for assigning them.

known applications of frequentist probability theory, and many useful problems outside the scope of frequentist probability theory, which are nevertheless solved readily by probability theory as logic.

Secondly, it is another widely held misconception that the mathematical rules of probability theory (the "laws of large numbers") would lead to (B) as a consequence of (A), and this seems to be "getting something for nothing." For, the fact that I know nothing about the coin is clearly not enough to make the coin give heads and tails equally often!

This misconception arises because of a failure to distinguish between the following two statements:

C: "Heads and tails are equally likely on a single toss."

D: "If the coin is tossed N times, each of the 2^N conceivable outcomes is equally likely."

To see the difference between (C) and (D), consider a case where it is known that the coin is biased, but not whether the bias favors heads or tails. Then (C) is applicable but (D) is not. For on this state of knowledge, as was noted already by Laplace, the sequences HH and TT are each somewhat more likely than HT or TH. More generally, our common sense tells us that any unknown influence which favors heads on one toss will likely favor heads on the other toss. Unless our robot has positive knowledge (symmetry of both the coin and the method of tossing) which definitely rules out *all* such possibilities, (D) is not a correct description of his true state of knowledge; it assumes too much.

Statement (D) implies (C), but says a great deal more. (C) says, "I do not know enough about the situation to give me any help in predicting the result of the next throw," while (D) says, "I know that the coin is honest, *and* that it is being tossed in a way which favors neither face over the other, *and* that the method of tossing and the wear of the coin give no tendency for the result of one toss to influence the result of another."

Mathematically, the laws of large numbers require much more than (C) for their derivation. Indeed, if we agree that tossing a coin generates an exchangeable sequence (i.e., the probability that N tosses will yield heads at n specified trials depends only on N and n, not on the order of heads and tails), then application of the de Finetti theorem, as in Chapter 9, shows that the weak law of large numbers holds *only* when (D) can be justified. In this case, it is almost correct to say that the probability assigned to heads is equal to the frequency with which the coin gives heads; because, for any $\epsilon \to 0$, the probability that the observed frequency $f = (n/N)$ lies in the interval $(1/2 \pm \epsilon)$ tends to unity as $N \to \infty$. Let us describe this by saying that there exists a *strong connection* between probability and frequency. We analyze this more deeply in Chapter 18.

In most recent treatments of probability theory, the writer is concerned with situations where a strong connection between probability and frequency is taken for granted – indeed this is usually considered essential to the very notion of probability. Nevertheless, the existence of such a strong connection is clearly only an ideal limiting case unlikely to be realized in any real application. For this reason, the laws of large numbers and limit theorems of probability theory can be grossly misleading to a scientist or engineer who naïvely supposes them to be experimental facts, and tries to interpret them literally in his problems. Here are two simple examples:

(1) Suppose there is some random experiment in which you assign a probability p for some particular outcome A. It is important to estimate accurately the fraction f of times A will be true in the next million trials. If you try to use the laws of large numbers, it will tell you various things about f; for example, that it is quite likely to differ from p by less than a tenth of one percent, and enormously unlikely to differ from p by more than one percent. But now, imagine that in the first hundred trials, the observed frequency of A turned out to be entirely different from p. Would this lead you to suspect that something was wrong, and revise your probability assignment for the 101'st trial? If it would, then your state of knowledge is different from that required for the validity of the law of large numbers. You are not sure of the independence of

different trials, and/or you are not sure of the correctness of the numerical value of p. Your prediction of f for a million trials is probably no more reliable than for a hundred.

(2) The common sense of a good experimental scientist tells him the same thing without any probability theory. Suppose someone is measuring the velocity of light. After making allowances for the known systematic errors, he could calculate a probability distribution for the various other errors, based on the noise level in his electronics, vibration amplitudes, etc. At this point, a naïve application of the law of large numbers might lead him to think that he can add three significant figures to his measurement merely by repeating it a million times and averaging the results. But, of course, what he would actually do is to repeat some unknown systematic error a million times. It is idle to repeat a physical measurement an enormous number of times in the hope that "good statistics" will average out your errors, because we cannot know the full systematic error. This is the old "Emperor of China" fallacy, discussed elsewhere.

Indeed, unless we know that all sources of systematic error – recognized or unrecognized – contribute less than about one–third the total error, we cannot be sure that the average of a million measurements is any more reliable than the average of ten. Our time is much better spent in designing a new experiment which will give a lower probable error *per trial.* As Poincaré put it, "The physicist is persuaded that one good measurement is worth many bad ones." In other words, the common sense of a scientist tells him that the probabilities he assigns to various errors do not have a strong connection with frequencies, and that methods of inference which presuppose such a connection could be disastrously misleading in his problems.

Then in advanced applications, it will behoove us to consider: How are our final conclusions altered if we depart from the universal custom of orthodox statistics, and relax the assumption of strong connections? Harold Jeffreys showed a very easy way to answer this, as we shall see later. As common sense tells us it must be, the ultimate accuracy of our conclusions is then determined not by anything in the data or in the orthodox picture of things; but rather by our own state of knowledge about the systematic errors. Of course, the orthodoxian will protest that, "We understand this perfectly well; and in our analysis we assume that systematic errors have been located and eliminated." But he does not tell us how to do this, or what to do if – as is the case in virtually every real experiment – they are unknown and so cannot be eliminated. Then all the usual 'asymptotic' rules are qualitatively wrong, and only probability theory as logic can give defensible conclusions.

The Arrogance of the Uninformed

Now we come to a very subtle and important point, which has caused trouble from the start in the use of probability theory. Many of the objections to Laplace's viewpoint which you find in the literature can be traced to the author's failure to recognize it. Suppose we do not know whether a coin is honest, and we fail to notice that this state of ignorance allows the possibility of unknown influences which would tend to favor the same face on all tosses. We say "Well, I don't see any reason why any one of the 2^N outcomes in N tosses should be more likely than any other, so I'll assign uniform probabilities by the principle of indifference."

We would be led to statement (D) and the resulting strong connection between probability and frequency. But this is absurd – in this state of uncertainty, we could not possibly make reliable predictions of the frequency of heads. Statement (D), which is supposed to represent a great deal of positive knowledge about the coin and the method of tossing can also result from *failure* to make proper use of all the available information!

Nothing in our past experience could have prepared us for this; it is a situation without parallel in any other field. In other applications of mathematics, if we fail to use all of the relevant data of a problem, the result will not be that we get an incorrect answer. The result will be that we are unable to get any answer at all. But probability theory cannot have any such built–in safety device, because in principle, the theory must be able to operate no matter what our incomplete information might be.

If we fail to include all of the relevant data or to take into account all the possibilities allowed by the data, probability theory will still give us a definite answer; and that answer will be the correct conclusion from the information that we actually gave the robot. But that answer may be in violent contradiction to our common–sense judgments which did take everything into account, if only crudely. *The onus is always on the user to make sure that all the information, which his common sense tells him is relevant to the problem, is actually incorporated into the equations and that the full extent of his ignorance is also properly represented.* If you fail to do this, then you should not blame Bayes and Laplace for your nonsensical answers.

We shall see examples of this kind of misuse of probability theory later, in the various objections to the Rule of Succession. It may seem paradoxical that a more careful analysis of a problem may lead to less certainty in prediction of the frequency of heads. However, look at it this way. It is commonplace that in all kinds of questions the fool feels a certainty that is denied to the wise man. The semiliterate on the next bar stool will tell you with absolute, arrogant assurance just how to solve all the world's problems; while the scholar who has spent a lifetime studying their causes is not at all sure how to do this.

In almost any example of inference, a more careful study of the situation, uncovering new facts, can lead us to feel either more certain or less certain about our conclusions, depending on what we have learned. New facts may support our previous conclusions, or they may refute them; we saw some of the subtleties of this in Chapter 5. If our mathematical model failed to reproduce this phenomenon, it could not be an adequate "calculus of inductive reasoning."