Appendix A

Other approaches to probability theory

Needless to say, the way we developed probability theory in Chapter 2 is not the only way it could have been done. The particular conditions we used might have been chosen differently, and there are several other approaches based on entirely different notions.

As an example of the former, many qualitative statements seem so obvious that one might think of taking them as basic axioms or desiderata, instead of the ones we did use. Thus if A implies B, then for all C we should expect intuitively to have $P(A|C) \le P(B|C)$. Of course, our rules do have this property, for the product rule is

$$P(AB|C) = P(B|AC)P(A|C) = P(A|BC)P(B|C). \tag{A.1}$$

But if A implies B, then $P(B|AC) = 1$ and $P(A|BC) \le 1$, so the product rule reduces to the intuitive statement. It may well be that a different choice of axioms would have simplified the derivations of Chapter 2. However, that was not the criterion we used. We chose the ones that appeared to us the most primitive and most difficult to quarrel with, in the belief that the resulting theory would be seen thereby to have the greatest possible generality and range of application.

Now we examine briefly some other approaches that have been advocated in the past.

A.1 The Kolmogorov system of probability

In our comments at the end of Chapter 2, we noted the Venn diagram and the relation to set theory that it suggests, which became the basis of the Kolmogorov approach to probability theory. This approach could hardly be more different from ours in general viewpoint and motivation; yet the final results are identical in several respects.

The Kolmogorov system of probability (henceforth denoted by KSP) is a game played on a sample space Ω of elementary propositions ω_i (or 'events'; it does not matter what we call them at this level of abstraction). We may think of them as corresponding roughly to the individual points of the Venn diagram, although of course the abstract definition makes no such reference.

Then there is a field F consisting of certain selected subsets f_j of Ω, corresponding roughly to our propositions $A, \ldots, B, \ldots$ represented by areas of the Venn diagram (although, again, the abstract definition allows sets which need not correspond to areas). F is to have basically three properties:

(I) Ω is in F;

(II) F is a σ-field, meaning that if f_j is in F, then its complement with respect to Ω, $\bar{f}_j = \Omega - f_j$, is also in F;

(III) F is closed under countable unions, meaning that, if countably many f_j are in F, their union is also in F.

Finally, there is to be a probability measure P on F, with the properties of:

(1) normalization: $P(\Omega) = 1$;
(2) non-negativity: $P(f_i) \geq 0$ for all f_i in F;
(3) additivity: if $\{f_1 \cdots f_n\}$ are disjoint elements of F (i.e. they have no points ω_i in common, then $P(f) = \sum_i P(f_j)$, where $f = \cup_j f_j$ is their union;
(4) continuity at zero: if a sequence $f_1 \supseteq f_2 \supseteq f_3 \supseteq \ldots$ tends to the empty set, then $P(f_j) \to 0$.

There is nothing surprising in these axioms; they seem to be familiar echoes of just what we found in Chapter 2, except that they state analogous properties of sets rather than propositions.

We are with Kolmogorov in spirit when he wants F to be a σ-field, for any proposition that can be affirmed can also be denied; the operation NOT was also one of our primitive ones. Indeed, we went further by including the operation AND. Then it was a pleasant surprise that (AND, NOT), which are an adequate set for deductive logic, turn out to be also adequate for our extended logic (i.e. given a set of propositions $\{A_1, \ldots, A_n\}$ to be considered, our rules generate a system of inference that is formally complete in the sense that it is adequate to assign consistent probabilities to all propositions in the Boolean algebra generated from $\{A_1, \ldots, A_n\}$.)

Kolmogorov's closure under countable unions is also implied by this requirement, in the following sense. Working fundamentally with finite sets, we are content fundamentally with finite unions; yet a well-behaved limit to an infinite set may be a convenient simplification, removing intricate but irrelevant details from a finite-set calculation. On the infinite sets thus generated, our finite unions go into countable unions.

But, as noted in Chapter 2, a proposition A referring to the real world cannot always be viewed as a disjunction of elementary propositions ω_i from any set Ω that has meaning in the context of our problem; and its denial $\overline{A}$ may be even harder to interpret as set complementation. The attempt to replace logical operations on the propositions $A, B, \ldots$ by set operations on the set Ω does not change the abstract structure of the theory, but it makes it less general in respects that can matter in applications. Therefore we have sought to formulate probability theory in the wider sense of an extension of Aristotelian logic.

Finally, the properties (1)–(4) of the probability measure P were stated by Kolmogorov as seemingly arbitrary axioms; and KSP has been criticized for that arbitrariness. But we recognize them as statements, in the context of sets, of just the four properties that we derived in Chapter 2 from requirements of consistency. For example, the need for non-negativity is apparent from (2.35). Additivity also seems arbitrary when stated merely as an axiom; but in (2.85) we have derived it as necessary for consistency.

Many writers have thought that normalization is merely an arbitrary convention, but (2.31) shows that if certainty is not represented by $p = 1$, then we must restate the sum and product rules, or we shall have an inconsistency. For example, if we choose the convention that $p = 100$ is to represent certainty, then these rules take the form

$$p(A|B) + p(\overline{A}|B) = 100, \qquad p(A|BC)p(B|C) = 100p(AB|C). \tag{A.2}$$

More generally, by any change of variables $u = f(p)$ with some monotonic function $f(p)$, we can represent probability on a different scale than the one adopted; but then consistency will require that the product and sum rules also be modified in form, so that the content of our theory is not changed. For example, with the change of variables $u = \log[p/(1-p)]$ the sum rule takes the equally simple form

$$u(A|B) + u(\overline{A}|B) = 0, \tag{A.3}$$

while the product rule becomes quite complicated. The substantive result is not that one is obliged to use any particular scale; but rather that a theory of probability whose *content* differs from one in which there is a single scale that is normalized, non-negative, and additive, will contain inconsistencies.

This should answer an objection sometimes raised (Fine, 1973, p. 65) that Kolmogorov's scale was arbitrary. In 1973, such a charge might have seemed reasonable, calling for further investigation. That further investigation has been made; since we now know that, in fact, he made the only choices that will pass all our tests for consistency, the charge now seems to us unjustified.

We do not know how Kolmogorov was able to see the need for his axiom (4) of continuity at zero; but our approach, in effect, derives it from a simple requirement of consistency. Firstly, let us dispel a misunderstanding. Statement (4) in terms of sets seems to imply that an infinite sequence of subsets is given. But its translation into a statement about propositions does not require that we assign probabilities to an infinite number of propositions. What is essential is that we have an infinite sequence of different *states of knowledge*, which may be about a single proposition, but which tends to impossibility. Since Kolmogorov's sets are not associated with any such idea as a 'state of knowledge', there seems to be no way to say this in the context of sets; but in the context of propositions it is easy.

We note this to emphasize that it would be a serious error to suppose that we can dispense with this axiom merely by limiting our discourse to a finite set of propositions. The resulting theory would have an arbitrary character which allows one to commit all kinds of inconsistencies.

In our system, 'continuity at zero' takes the following form: given a sequence of probabilities $p(A)_1, p(A)_2, \ldots$ that tend to certainty, the probabilities $p(\overline{A})_1,\ p(\overline{A})_2, \ldots$ assigned to the denial must tend to zero. Indeed, as we noted in (2.48), the functional equation that $S(x)$ satisfies ties values at different x together so strongly that the exact way in which $S(x)$ tends to zero as $x \to 1$ is the crucial thing that determines the function $S(x)$ over its entire range $(0 \leq x \leq 1)$, and therefore determines the additivity property (2.58). Thus, from our viewpoint, Kolmogorov's axioms (3) and (4) appear to be closely related; it is not obvious whether they are logically independent.

For all practical purposes, then, our system will agree with KSP if we are applying it in the set-theory context. But in more general applications, although we have a field of discourse F and probability measure P on F with the same properties, we do not need, and do not always have, any set Ω of elementary propositions into which the elements of F can be resolved. Of course, in many of our applications such a set Ω will be present: for example, in equilibrium statistical mechanics the elements ω_i of Ω can be identified with the stationary 'global' quantum states of a system, which comprise a countable set. In these cases, there will be essentially complete agreement in the abstract formulation, although we carry out practical calculations with more freedom in one respect – and more inhibition in another – for reasons noted below.

Our approach supports KSP in another way also. KSP has been criticized as lacking connection to the real world; it has seemed to some that its axioms are deficient because they contain no statement to the effect that the measure P is to be interpreted as a frequency in a random experiment.[1] But, from our viewpoint, this appears as a merit rather than a defect; to require that we invoke some random experiment before using probability theory would have imposed an intolerable and arbitrary restriction on the scope of the theory, making it inapplicable to most of the problems that we propose to solve by extended logic.

Even when random experiments are involved in the real problem, propositions specifying frequencies are properly considered, not as determinations of the measure P, but as elements of the field F. In both Kolmogorov's system and ours, such propositions are not the tools for making inferences, but the things about which inferences are being made.

There are some important differences, however, between these two systems of probability theory. In the first place, in KSP attention is concentrated almost exclusively on the notion of additive

[1] Indeed, de Finetti (1972, p. 89) argues that Kolmogorov's system *cannot* be interpreted in terms of limits of frequencies.

measure. The Kolmogorov axioms make no reference to the notion of conditional probability; indeed, KSP finds this an awkward notion, really unwanted, and mentions it only reluctantly, as a seeming afterthought.[2] Although Kolmogorov has a section entitled 'Bayes' theorem', most of his followers ignore it. In contrast, we considered it obvious from the start that all probabilities referring to the real world are necessarily conditional on the information at hand. In Chapter 2 the product rule, with conditional probability and Bayes' theorem as immediate consequences, appeared in our system even before additivity.

Our derivation showed that, from the standpoint of logic, the product rule (and therefore Bayes' theorem) expresses simply the associativity and commutativity of Boolean algebra. This is what gives us that greater freedom of action in calculations, leading in later chapters to the unrestricted use of Bayes' theorem, in which we have complete freedom to move propositions back and forth between the left and right sides of our probability symbols in any way permitted by the product and sum rules. This is a superb computational device – and by far the most powerful tool of scientific inference – yet it is completely missing from expositions of probability theory based on the KSP work (which do not associate probability theory with information or inference at all).

In return for this freedom, we impose on ourselves an inhibition not present in KSP. Having been burned by de Finetti and his followers, we are wary of infinite sets, and approach them only cautiously, after ascertaining that in our problem there is a well-defined and well-behaved limiting process that will not lead us into paradoxes and will serve a useful purpose.

In principle, we start always by enumerating some finite set of propositions $A, B, \dots$ to be considered. Our field of discourse F is then also finite, consisting of these and – automatically – all propositions that can be 'built up' from them by conjunction, disjunction, and negation. We have no need or wish to 'tear down' by resolving them into a disjunction of finer propositions, much less carrying this to infinite limits, except when this can be a useful calculational device due to the structure of a particular problem.

We have three reasons for taking this stance. The first was illustrated in Chapter 8 by the scenario of Sam's broken thermometer, where we saw that beyond a certain point this finer and finer resolving serves no purpose. Secondly, in Chapter 15 we saw some of the paradoxes that await those who jump directly into infinite sets without considering any limiting process from a finite set. But even here, when we considered the so-called 'Borel–Kolmogorov paradox', we found ourselves in agreement with Kolmogorov's resolution of it, and thus in disagreement with some of his critics. One must approach infinite sets carefully; but once in an uncountable set, one must then approach sets of measure zero just as carefully.

A third reason is that a different resolution often appears more useful to us. Instead of resolving a proposition A into the disjunction $A = B_1 + B_2 + \cdots$ of 'smaller' propositions and applying the sum rule, one can equally well resolve it into a conjunction $A = C_1 C_2 \cdots$ of 'larger' propositions and apply the product rule. This may be interpreted, in terms of sets, very simply. To specify the geographical layout of a country, there are two possible methods: (1) specify the points that are in it; (2) specify its boundary. Method (1) is the Venn–Kolmogorov viewpoint; but method (2) appears to us equally fundamental and often simpler and more directly related to the information we have in a real problem. In a Venn diagram the boundary of set A is composed of segments of the boundaries of $C_1, C_2, \dots$, just as that of a country is composed of segments of rivers, coastlines, and adjacent countries.

These methods are not in conflict; rather, in each problem we may choose the one appropriate to the job before us. But in most of our problems method (2) is the natural one. A physical theory is always stated as a conjunction of hypotheses, not a disjunction; likewise a mathematical theory is defined by the set of axioms underlying it, which is always stated as a conjunction of elementary

[2] In the Kolmogorov system, conditional probability is such a foreign element that an entire book has been written (Rao, 1993) trying to explain the idea of conditional probability by giving it a separate axiomatic approach!

axioms. To express the foundations of any theory as disjunctions would be almost impossible; so we must demand this freedom of choice.

In summary, we see no substantive conflict between our system of probability and Kolmogorov's as far as it goes; rather, we have sought a deeper conceptual foundation which allows it to be extended to a wider class of applications, required by current problems of science.

The theory expounded here is still far from its final, complete form, however. In its present state of development, there are many situations where the robot does not know what to do with its information. For example, suppose it is told that 'Jones was very pleased at the suggestion that θ might be greater than 100'. By what principles is it to translate this into a probability statement about θ?

You and I, however, can make some use of that information to modify our opinions about θ (upward or downward according to our opinions about Jones). Indeed we can use almost any kind of prior information, and perhaps draw a free-hand curve which indicates roughly how it affects our probability distribution $p(\theta)$. In other words, *our brains are in possession of more principles than the robot's for converting raw information, semiquantitatively, into something which the computer can use*. This is the main reason why we are convinced that there must be more principles like maximum entropy and transformation groups, waiting to be discovered by someone. Each such discovery will open up a new area of useful applications of this theory.

A.2 The de Finetti system of probability

There is today an active school of thought, most of whose members call themselves 'Bayesians', but who are actually followers of Bruno de Finetti and concerned with matters that Bayes never dreamt of. In 1937, de Finetti published a work which expressed a philosophy somewhat like ours and contained not only his marvellous and indispensable exchangeability theorem, but also sought to establish the foundations of probability theory itself on the notion of 'coherence'. This means, roughly speaking, that one should assign and manipulate probabilities so that one cannot be made a sure loser in betting based on them. He appears to derive the rules of probability theory very easily from this premise (de Finetti, 1937).

Since 1937, de Finetti has published many more works on this topic (de Finetti, 1958, 1974a,b), as cited in our general references. Note particularly the large work published in English translation (de Finetti, 1974b). Some have thought that we should have followed de Finetti's coherence principle in the present work. Certainly, that would have shortened our derivations. However, we think that coherence is an unsatisfactory basis in three respects. The first is admittedly only aesthetic; it seems to us inelegant to base the principles of logic on such a vulgar thing as expectation of profit.

The second reason is strategic. If probabilities are thought to be defined basically in terms of betting preferences, then for assigning probabilities one's attention is focused on how to elicit the personal probabilities of different people. In our view, that is a worthy endeavor, but one that belongs to the field of psychology rather than probability theory; our robot does not have any betting preferences. When we apply probability theory as the normative extension of logic, our concern is not with the personal probabilities that different people might happen to have, but with the probabilities that they 'ought to' have, in view of their information – just as James Clerk Maxwell noted in our opening quotation for Chapter 1.

In other words, at the beginning of a problem our concern is not with anybody's personal opinions, but with specifying the *prior information* on which our robot's opinions are to be based, in the context of the current problem. The principles for assigning prior probabilities consistently by logical analysis of that prior information are for us an essential part of probability theory. Such considerations are almost entirely absent from expositions of probability theory based on the de Finetti approach (although of course it does not forbid us to consider such problems).

The third reason is thoroughly pragmatic: if any rules were found to possess the property of coherence in the sense of de Finetti, but not the property of consistency in the sense of Cox, they would be clearly unacceptable – indeed, functionally unusable – as rules for logical inference. There would be no 'right way' to do any calculation, and no 'right answer' to any question. Then there would be small comfort in the thought that all those different answers were at least 'coherent'.

To the best of our knowledge, de Finetti does not mention consistency as a desideratum, or test for it. Yet it is consistency – not merely coherence – that is essential here, and we find that, when our rules have been made to satisfy the consistency requirements, then they have automatically (and trivially) the property of coherence.

Another point was noted in our Preface. Like Kolmogorov, de Finetti is occupied mostly with probabilities defined directly on arbitrary uncountable sets; but he views additivity differently, and is led to such anomalies as an unlimited sequence of layers, like an onion, of different orders of zero probabilities that add up to one, etc. It is the followers of de Finetti who have perpetrated most of the infinite-set paradoxing that has forced us to turn to (and, like Helmholtz in Chapter 16, exaggerate if necessary) the opposite 'finite-sets' policy in order to avoid them. This line of thought continues, with technical details, in Chapter 15.

A.3 Comparative probability

In our Comments Section 1.8 at the end of Chapter 1, we noted a possible objection to our first desideratum:

(I) Degrees of plausibility are to be represented by real numbers.

Why must one do this? Our pragmatic reason was that we do not see how our robot's brain can function by carrying out definite physical operations – mechanical or electronic – unless, at some point, degrees of plausibility are associated with some definite physical quantity.

We recognize that this ignores some aesthetic considerations; for example, the geometry of Euclid derives its elegance in large part from the fact that it is not concerned with numerical values, but with recognizing qualitative conditions of equivalence or similarity. We had this very much in mind when choosing all our other axioms, being careful to ensure that 'consistency' and 'correspondence with common sense' expressed qualitative rather than quantitative properties.

Of course, our one pragmatic argument carries no weight for those concerned with abstract axiomatics rather than making something work, so let us consider the alternatives. If one wishes to pick away at our desideratum I, it can be dissected into simpler axioms. In the following, read '$(A|C) > (B|C)$' not as a numerical comparison, but simply as the verbal statement, 'given C, A is more plausible than B', etc. Then desideratum I may be replaced by two more elementary ones:

(Ia) Transitivity. If $(A|X) \geq (B|X)$ and $(B|X) \geq (C|X)$ then $(A|X) \geq (C|X)$.
(Ib) Universal comparability. Given propositions A, B, C, then one of the relations $(A|C) > (B|C)$, $(A|C) = (B|C)$, $(A|C) < (B|C)$ must hold.

To see this, note that, if we postulate both transitivity and universal comparability, then within any finite set of propositions, we can always set up a representation by real numbers (in fact, by rational numbers) that obeys all the ordering relations. For, suppose we have a set $\{A_1, \dots, A_n\}$ of propositions with such numerical measures $\{x_1, \dots, x_n\}$. Adding a new proposition A_{n+1}, the transitivity and universal comparability ensure that it fits into a unique place in those ordering relations. But since between any two rational numbers one can always find another rational number, we can always assign a number x_{n+1} to it so that all the ordering relations of the new set $\{A_1, \dots, A_{n+1}\}$ are also obeyed by our rational numbers $\{x_1, \dots, x_{n+1}\}$.

At this point, therefore, it is all over with any comparative theory which embodies both transitivity and universal comparability. Once the existence of a representation by real numbers is established, then Cox's theorems take over and force that theory to be identical with the theory of inference that we derived in Chapter 2. That is, either there is some monotonic function of the x_i that obeys the standard product and sum rules of probability theory; or else we can exhibit inconsistencies in the rules of the comparative theory.

Some systems of comparative probability theory have both of these axioms; then they have assumed everything needed to guarantee the equivalence to the standard numerical valued theory. This being the case, it would seem foolish to refuse to use the great convenience of the numerical representation. But now, can one drop transitivity or universal comparability and get an acceptable extension of logic with different content than ours?

No comparative probability theory is going to get far if it violates transitivity. Nobody would wish to – or be able to – use it, because we would be trapped in endless loops of circular reasoning. So transitivity is surely going to be one of the axioms of a comparative probability theory; discovery of an intransitivity would be grounds for immediate rejection of any system.

To many, universal comparability does not seem a compelling desideratum. By dropping it we could create a 'lattice' theory, so called because we can represent propositions by points, relations of comparability by lines connecting them in various ways. Then it is conceivable that A and C can be compared, and B and C can be compared; but A is not comparable to B. One might contemplate a situation in which $(A|D) < (C|D)$ and $(B|D) < (C|D)$; but neither $(A|D) < (B|D)$ nor $(A|D) \geq (B|D)$ could be established. This allows a looser structure which cannot be represented faithfully by assigning a single real number to each proposition (although it can be so represented by a lattice of vectors); any attempt to introduce a single-valued numerical representation would generate false comparisons not present in the system.

Much effort has been expended on attempts to develop such looser forms of probability theory in which one does not represent degrees of plausibility by real numbers, but admits only qualitative ordering relations of the form $(A|C) \geq (B|C)$, and attempts to deduce the existence of a (not necessarily unique) additive measure $p(A|B)$ with the property (2.85). The work of L. J. Savage (1954) is perhaps the best known example. A summary of such attempts is given by T. L. Fine (1973). These efforts appear to have been motivated only by an aesthetic feeling – that universal comparability is a stronger axiom than we need – rather than from the hope that any particular pragmatic advantage would be gained by dropping it. However, a restriction appeared in comparative probability theory, robbing it of its initial appeal.

Ordering relations may not be assigned arbitrarily because it must always be possible to extend the field of discourse, by adding more propositions and ordering relations, without generating contradictions. If adding a new ordering relation created an intransitive loop, it would be necessary to modify some ordering relations to restore transitivity. But such extensions may be carried out indefinitely, and, when a set of propositions with transitive ordering relations becomes, in a certain sense, 'everywhere dense' on the path from impossibility to certainty, consistency will require that the theory then approach the conventional numerical-valued probability theory expounded here.

In retrospect (i.e. in view of Cox's consistency theorems) this is hardly surprising; a comparative probability theory whose results conflict with those of our numerical probability theory necessarily contains within it either overtly visible inconsistencies or the seeds of inconsistencies which will become visible when one tries to extend the field of discourse.

Furthermore, it appears to us that any computer designed to carry out the operations of a comparative probability theory must at some stage represent the ordering relations as inequalities of real numbers. So attempts to evade numerical representation not only offer no pragmatic advantage, they are futile. Thus in the end the study of comparative probability theories serves only to show us still another aspect of the superiority of the Cox approach that we follow here.

A.4 Holdouts against universal comparability

The arguments in the preceding sections, however, do not quite close the subject, because some of the criticisms of probability theory as logic are from writers who have considered it absurd to suppose that all propositions can be compared. This view seems to arise from two different beliefs: (1) human brains cannot do this; and (2) they think that they have produced examples where it is fundamentally impossible to compare all propositions.

Argument (1) carries no weight for us; in our view, human brains do many absurd things while failing to do many sensible things. Our purpose in developing a formal theory of inference is not to imitate them, but to correct them. We agree that human brains have difficulty in comparing, and reasoning from, propositions that refer to different contexts. But we would observe also that the ability to do this improves with education.

For example, is it more likely that (A) Tokyo will have a severe earthquake on June 1, 2230; or that (B) Norway will have an unusually good fish catch on that day? To most people, the contexts of propositions A and B seem so different that we do not see how to answer this. But with a little education in geophysics and astrophysics, one realizes that the moon could well affect both phenomena, by causing phase-locked periodic variations in the amplitudes of both the tides and stresses in the Earth's crust. Recognition of a possible common physical cause at work makes the propositions seem comparable after all.

The second objection to universal comparability noted above appears to be a misunderstanding of our present theory, but one which does point to cases in which universal comparability would indeed be fundamentally impossible. These are the cases where we are trying to classify propositions with respect to more than one attribute, as in the conceivable multidimensional models of mental activity noted at the end of Chapter 1. All of the alleged counter-examples to comparability that we have seen prove on examination to be of this type.

For example, a mineralogist may classify a collection of rocks with respect to two qualities, such as density and hardness. If, within a certain subclass of them, density alone varies, then obviously there are transitive comparability relations that can be represented faithfully by real numbers d. If in another subclass hardness alone varies, there is a similar comparability representable by real numbers h. But if we classify rocks by both simultaneously, it requires two real numbers $(d,\ h)$ to represent them; any attempt to arrange them in a unique one-dimensional order would be arbitrary.

The arbitrariness could be removed if we also introduced some new value judgment or 'objective function' $f(d, h)$ that tells us by relations such as $f(d_1, h_1) = f(d_2, h_2)$ how to trade off a change $\Delta\, d = d_2 - d_1$ in d against a change $\Delta\, h = h_2 - h_1$ in h. But then we are classifying the rocks with respect to only one attribute, namely f, and universal comparability is again possible.

In the theory of probability developed here we are, by definition, classifying propositions according to only one attribute, which we call intuitively 'degree of plausibility'. Once this is understood, we think that the *possibility* of representation by real numbers need never be questioned, and the *desirability* of doing this is attested to by all the nice results and useful applications of the theory.

Nevertheless, the general idea of a comparative probability theory might be useful to us in two respects. Firstly, for many purposes one has no need for precisely defined numerical probabilities; any values that preserve ordering relations within a small set of propositions may be adequate for our purpose. For example, if it is required only to choose between two competing hypotheses, or two feasible actions, a wide range of numerical probability values must all lead to the same final choice. Then the precise position within that range is irrelevant, and to determine it would be wasted computational effort. Something much like a comparative probability theory would then appear, not as a generalization of numerical probability theory, but as a simple, useful approximation to it.

Secondly, the above observation about Tokyo and Norway suggests a possible legitimate application for a lattice theory of probability. If our brains do not have automatically the property of universal comparability, then perhaps a lattice theory might come much closer than the Laplace–Bayes theory to describing the way we actually think. What are some of the properties that can be anticipated of a lattice theory?

A.5 Speculations about lattice theories

One evident property is that we could do plausible reasoning only in certain 'domains' consisting of sets of comparable propositions. We would not have any idea how to reason in cases involving a jump across widely separated parts of the lattice; unless we perceive some logical relationship between propositions, we have no criterion for comparing their plausibilities. Our scale of plausibility might be wildly different on different parts of the lattice, and we would have no way of knowing this until we had learned to increase the degree of comparability.

Indeed, the human brain does not start out as an efficient reasoning machine, plausible or deductive. This is something which we require years to learn, and a person who is an expert in one field of knowledge may do only rather poor plausible reasoning in another.[3] What is happening in the brain during this learning process?

Education could be defined as the process of becoming aware of more and more propositions, and of more and more logical relationships between them. Then it seems natural to conjecture that a small child reasons on a lattice of very open structure: large parts of it are not interconnected at all. For example, the association of historical events with a time sequence is not automatic; the writer has had the experience of seeing a child, who knew about ancient Egypt and had studied pictures of the treasures from the tomb of Tutankhamen, nevertheless coming home from school with a puzzled expression and asking: 'Was Abraham Lincoln the first person?'

It had been explained to him that the Egyptian artifacts were over 3000 years old, and that Abraham Lincoln was alive 120 years ago; but the meaning of those statements had not registered in his mind. This makes us wonder whether there may be primitive cultures in which the adults have no conception of time as something extending beyond their own lives. If so, that fact might not have been discovered by anthropologists, just because it was so unexpected that they would not have raised the question.

As learning proceeds, the lattice develops more and more points (propositions) and interconnecting lines (relations of comparability), some of which will need to be modified for consistency in the light of later knowledge. By developing a lattice with denser and denser structure, *one is making his scale of plausibilities more rigidly defined*.

No adult ever comes anywhere near to the degree of education where he would perceive relationships between all possible propositions, but he can approach this condition with some narrow field of specialization. Within this field, there would be a 'quasi-universal comparability', and his plausible reasoning within this field would approximate that given by the Laplace–Bayes theory.

A brain might develop several isolated regions where the lattice was locally quite dense; for example, one might be very well-informed about both biochemistry and musicology. Then for reasoning within each separate region, the Laplace–Bayes theory would be well-approximated, but there would still be no way of relating different regions to each other.

[3] The biologist James D. Watson has remarked before TV cameras that professional physicists can be 'rather stupid' when they have to think about biology. We do not deny this, although we wonder how far he would have got in finding the DNA structure without the help of the physicists Rosalind Franklin, to acquire the data for him, and Francis Crick, to explain to him how to interpret it.

Then what would be the limiting case as the lattice becomes everywhere dense with truly universal comparability? Evidently, the lattice would then collapse into a line, and some unique association of *all* plausibilities with real numbers would then be possible. Thus, *the Laplace–Bayes theory does not describe the inductive reasoning of actual human brains; it describes the ideal limiting case of an 'infinitely educated' brain*. No wonder that we fail to see how to use it in all problems!

This speculation may easily turn out to be nothing but science fiction; yet we feel that it must contain at least a little bit of truth. As in all really fundamental questions, we must leave the final decision to the future.