Appendix

1 Concerning Various Aspects of the Different Approaches

In every field, and in particular in the calculus of probability, there is scope, both hypothetically and in fact, for a number of axiomatic approaches, each of which, to a greater or lesser degree, differs from the others in various respects. It does not suit our purpose to choose just one of these, merely illustrating – even if exhaustively – that particular one; nor are we interested in presenting a somewhat wide and diverse collection from which each person makes his choice with the aid of a pin. The way that seems more appropriate, and that in any case we shall try to follow, consists in sticking to one preferred approach as a reference point but at the same time illustrating both the variants within it that seem admissible, or necessary, and the approaches inspired by divergent views. This provides the framework for the necessary conceptual and formal comparisons.

From a conceptual standpoint our choice has already been made, and explained at some length, in Chapters 3, 4 and subsequently. At that time, we gave what might be called an axiomatic approach, but between then and now there is a difference in attitude that can be expressed (in the summary form of a single sentence) by saying that we must pass from an axiomatic approach to the *theory of probability*, to an axiomatic approach to the *calculus of probability*. This transition must not be taken as implying the existence of any distinction or separation between the two terms, or the desirability of creating such a distinction; we simply wish to draw attention to the different perspective that is obtained by emphasizing on the one hand the essential meaning, on the other the formal aspect.

The difference in perspective is the same as that which occurs when a given theory is viewed by a physicist and a mathematician. One concentrates his attention on the passage from the 'facts' to their mathematical translation; the other on the work involved in the latter. This then resolves itself into the difference between the axiomatization of a theory considered *from the point of view of its meaning*, and the axiomatization of a theory reduced *to its formal and abstract aspect*. In the first case, with reference to the example of a physical theory, the *axioms* encapsulate all those properties of an experimental nature that have been ascertained (or are assumed, perhaps hypothetically, to have been ascertained), and which suffice to give meaningful (i.e. operational) definitions of concepts and quantities, and to establish a mathematical theory to which they are subjected. In the second case, however, we omit the details and merely assume

Theory of Probability: A Critical Introductory Treatment, First Edition. Bruno de Finetti.
© 2017 John Wiley & Sons Ltd. Published 2017 by John Wiley & Sons Ltd.

the result as our starting point: the *axioms*, independently of the meaning and validity of the physical interpretation from which one starts, are now nothing more than an expression of the mathematical nature of certain entities and of the form of the relations among them. In this way, the mathematician can work with the axioms without worrying about those features which do not concern him *qua* mathematician. As always, the division of labour carries with it both advantages and disadvantages. A blind man with very acute hearing and a deaf man with very sharp eyesight will be able, in conjunction, to see and hear better than a normal individual, but they might 'understand' less owing to their inability to communicate. We will return to this point later.

The distinction we have just considered applies equally to the case of probability. As an axiomatic approach to the *theory of probability*, we understand the axiomatization made from the point of view of *meaning*. The latter consisted, for us, in the analysis of the conditions of coherence for bets (or something similar) on things we called 'events'; for others, it may consist of assertions either about symmetries, or frequencies, or things also called 'events', but which, perhaps, might be thought of as 'sequences of events', or whatever. In this way, one comes to impart meaning to certain words (quantities etc.) and to establish relationships that must hold among them. As an axiomatic approach to the *calculus of probability*, we mean the axiomatization made from the *formal and abstract* point of view: we have rules with which to operate on symbols without the necessity of knowing which, if any, interpretation these rules and symbols have in the actual context.

Of course, such a contraposition is too crude to serve as anything other than a starting point; on no account must we gloss over the finer points (perhaps hidden to a superficial view, but nonetheless essential). In the choice of the mathematical axiomatization there is plenty of scope for choosing among formulations that are formally equivalent (but whose particular axioms, to those who recall the original meaning, might differ in their intuitive appeal); on the other hand, choices that are made concerning the more 'peripheral' aspects can appear rather arbitrary and made simply for mathematical convenience.

The path we shall follow is motivated by our steadfast refusal to adopt this bad habit. In precise terms: *the axioms of the calculus of probability will be nothing more, and nothing less, than the translation into an abstract form of the conclusions which follow strictly from the practical exigencies brought to light during the preliminary discussions concerning the theory of probability*. It is useful, at this point, to clarify, in a summary and preliminary fashion, why this statement, so obvious in itself, is, instead, at odds with all those formulations, which, by following the same criterion *a little less strictly*, end up, in my opinion, by not following it at all. These clarifications will certainly not be enough to give a sufficient picture of the many factors to be taken into consideration and of their compass. However, they will enable those who bear them in mind to get to grips with the many considerations that will have to be worked out in detail, but without repeating too often, and tediously, these general motives.

We know that what we have to deal with in any case will be the characterization of certain functions **P** defined over the field of entities E, called 'events' (and then **P** is called 'probability'), or over the wider field of entities X called 'random quantities' (and then **P** is called 'prevision'). In order to carry out our task we will try to pose the formal questions concerning events in such a way as to reproduce, as faithfully as possible, the circumstances that can practically arise for events (together with variants – some important, some less so – to meet particular exigencies): similarly for random quantities. In order to define the functions acceptable as **P**, we will utilize only the conditions of coherence expressed in an abstract form.

In what way does this differ from the formulations more usually adopted at the present time? In the first place, the structure which is generally preferred is a closed, monolithic one. Rather than defining events in a general way, and then the functions **P** as extendible (in principle) to all events (either already conceived of, or conceivable in the future), one constructs on each occasion a definite, well-delimited (although possibly enormous) field of events with a particular function **P** attached to it once and for all. In terms of the standard image (in which events are thought of as sets in an abstract space), this means that the complete set-up (or 'probability space') is a *measure space* (i.e. a space with *one* particular, fixed measure). In contrast to this, the separate consideration of first the *space* (*without the measure*, or any other kind of structure) and then all the possible *measures*, not only, and most importantly, meets the needs of the subjective conception by providing $\mathbf{P}_i$, which are possibly different for each individual i ('tot capita, tot sententiae'), but also satisfies other more 'neutral' requirements (probabilities conditional on different hypotheses, or different states of information, or 'mixtures', and so on).

Moreover (independently of the previous objection, concerning **P**), this space–measure coupling gives rise to an unnatural, forced relationship between the two notions of event and probability, because it does not take account of the problems raised by the fixing of a particular function **P**. The current practice of reducing the calculus of probability to modern measure theory (countably or σ-additive, as in the Lebesgue theory) – apart from changes in terminology (set–event; measure–probability; function–random quantity; integral–expectation) – has resulted in the following:

- probability is obliged to be not merely additive (as is necessary) but, in fact, σ-additive (without any good reason);
- events are restricted to be merely a subclass (technically, a σ-ring with some further conditions) of the class of all subsets of the space (in order to make σ-additivity possible, but without any real reason that could justify saying to one set 'you are an event', and to another 'you are not');
- people are led to extend the set of events in a fictitious manner (i.e. not corresponding to any meaningful interpretation) in order to preserve the appearance of σ-additivity even when it does not hold (in the meaningful field), rather than abandoning it.

Among other things, in the case of limiting processes and definitions of stochastic limits, this leads to the adoption of formulations that are unacceptable as they stand unless σ-additivity is imposed (at the cost of a great deal of inconvenience) as a necessary assumption at all times.

We should, of course, discuss these objections and reservations rather more fully, and go on to justify them; all the more so in that they will seem strange to those who are accustomed to the standard formulation. In fact, in the standard approach the points which do not seem to stand up to a critical examination are introduced either with the tacit suggestion that they are obvious, or they are couched in suitably seductive terms to overcome any initial reluctance to accept them.

There are other negative features of the space–measure coupling which are not related to the assumption of σ-additivity. An example is provided by the fact that *zero probability* is regarded as a property of the event in question (among other things, this sometimes leads to two events, $A \neq B$, being defined as 'equivalent' if their symmetric difference, $A\tilde{B}+B\tilde{A}$, or $A + B - 2AB$, has zero probability). Even more dangerous is the fact that *stochastic independence* – $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$ – is considered as being a property of the events; and so on. One should beware of laying insufficient stress on the fact that it is a property of the function **P** (in relation to the events A and B) and not of the events as such (but this important distinction ceases to have any meaning if **P** is considered as given!).

In Chapter 2, we gave an account of what can be said about events from a logical standpoint (within the logic of certainty); in other words, concerning the events in themselves. We postponed until Chapter 3 anything which depended on the introduction of the function **P** defined on the events (without adding or altering anything concerning the notion of event, or the meanings of individual events). This separation was made in order to avoid any confusion early on; confusion which could have led to misunderstandings later.

Here we shall adopt the same policy, although, of course, in a deeper, more systematic and precise way. Certain distinctions that appear meaningful in other formulations no longer appear so in ours. Consider, for example, the distinctions between whether or not events are *atomic* (i.e. contain no events other than themselves and the empty one; in terms of sets of points, this reduces to those sets formed from the singletons), or between those events *belonging to either finite or infinite sets* of events, and so on. In the case of a random quantity X, having as possible values, for example, all real numbers between 0 and 1 (like X = 'percentage of time during which a given telephone link will be busy tomorrow between 9 a.m. and 5 p.m.'), let us consider the event $E = (X = 0{\cdot}4166666\ldots)$. It consists of obtaining exactly a given preassigned value and could be regarded both as 'belonging to an infinite set' (i.e. of events $E_x = (X = x)$, $0 \leqslant x \leqslant 1$) and as an 'atomic event' (because a precise value, like $x_0 = 41\frac{2}{3}\%$, does not admit further refinement). It also belongs, however, to the field consisting of just the two events $E = (X = x_0)$ and $\tilde{E} = (X \neq x_0)$ (together with the events 0 and 1), and can be decomposed into $E = EA + E\tilde{A}$, by means of any event A not involving X (for example: A = 'it will rain tomorrow'; A = 'the party at present in government will not remain in power after the next election'; A = 'the azaleas in the window of the florist across the street will be sold today'). This can be extended to infinite subcases by considering other random quantities $Y, Z,\ldots$ (and, therefore, by considering as 'provisional atoms' the points (x, y), or (x, y, z), or $(x, y, z, \ldots)$ of $S_2, S_3,\ldots, S_n,\ldots$). It follows that any considerations put forward on the basis of these nonexistent distinctions must be without foundation (an example of this is the assertion that an event E that is not impossible can only have zero probability, $\mathbf{P}(E) = 0$, if it 'belongs to an infinite set of events').

On the other hand, there exist real problems that arise in various connections with the notion of the 'verifiability' of an event; a notion which is often vague and elusive. Strictly speaking, the phrase itself is an unfortunate one because verifiability is the essential characteristic of the definition of an event (to speak of an 'unverifiable event' is like saying 'bald with long hair'). It is necessary, however, to recognize that there are various degrees and shades of meaning attached to the notion of verifiability. Some are more or less flexible: verifiable with a greater or lesser degree of *precision*; or within a shorter or longer period of *time*; or with a higher or lower level of *expenditure*; or with a greater or lesser *number* of partial verifications; and so on. Others are more precise: for example, we could consider 'absolute' degrees of precision, or 'infinite' time periods, and so on. The most precise and important, however, is that which arises in theoretical physics in connection with *observability* and *complementarity*. It seems strange that a question of such overwhelming interest, both conceptually and practically (and concerning the most unexpected and deep forms of application of probability theory to the natural sciences), should be considered, by and large, only by physicists and philosophers, whereas it is virtually ignored in treatments of the calculus of probability. We agree that it is a new element, whose introduction upsets the existing framework, making it something of a hybrid. We see no reason, however, to prefer tinkering about with bogus innovations rather than enriching the existing structure by incorporating stimulating refinements (disruptive though they may be).

It is our intention, therefore, to attempt to provide in this appendix an integrated view of questions of this kind that arise in connection with events. We should perhaps make it clear that our 'attempt' will be mainly concerned with the case of theoretical physics, and will consist of little more than a comparison of the positions adopted by various other authors, plus an indication of which position seems to us to be less open to criticism (as well as being better suited to deal with further problems concerning the verifiability of events).

There are other questions (already mentioned many times in passing) which concern the notion of 'possibility', and further aspects are revealed in cases where, either through haste or oversight (or because of one's own limitations, or because of the impossible nature of the task, or whatever), one has not drawn out all the logical implications contained[1] in the information in one's possession. The result of this is that the set of events considered 'certain' is not *closed* with respect to the logic of certainty.

Finally, we shall turn from the preliminary questions concerning events (and hence the logic of certainty) to the introduction of probability. It is the latter that is for us the real subject matter, the principle character as it were, and the rest is simply the setting of the scene.

We must now pass from the considerations that led us (in Chapter 3) to our basic formulation, to consider the *axioms* which constitute their translation into abstract form. The surest way of avoiding any kind of modification taking place during this translation is to directly express things in abstract form without any alterations. It suffices to preserve additivity and non-negativity. So far as the essential considerations are concerned, this rules out the attributing of a positive price (positive prevision) to a transaction (or bet) that will certainly lead to a negative outcome. From the abstract point of view, this obliges $\mathbf{P}$ to be such that we can never have

$$c_1\mathbf{P}(X_1)+c_2\mathbf{P}(X_2)+\ldots+c_n\mathbf{P}(X_n)>0$$

if

$$X=c_1X_1+c_2X_2+\ldots+c_nX_n \text{ is certainly} <0.$$

These inequalities (imposed for every finite, linear combination) define (as the intersection of half-spaces) the convex set P of admissible functions $\mathbf{P}$ (and all that remains to be sorted out are a few details, like the possibility of substituting $\geqslant$ for > in the inequalities, and so on).

1 As an example of this, consider the matching problem (with n objects). It could happen that someone does not realize that the case of $n-1$ matchings is impossible (see Chapter 3, 3.8.4), either because he is not capable of arriving at this conclusion on the basis of the information available to him, or because it never occurred to him to doubt that all the values 0 to n were possible. It could also be that he had once known the result, but had subsequently forgotten it; or that he had not really forgotten it, but simply overlooked it at the time in question.

On the other hand, we should point out that in expressing these conditions we have made use of, or at least made reference to, random quantities rather than events. In actual fact, writing $E_1 \ldots E_n$ in place of $X_1 \ldots X_n$ would have given practically the same condition[2] by introducing the X_h (which form a linear space) in an indirect fashion as linear combinations of the events (which do not form a linear space). To start directly with the linear space of the X_h (without giving any particular emphasis to events, which are, in any case, part of that space) not only enables one to deal with the whole set-up in one go but also permits one to emphasize the adherence to the essential meaning.

Proceeding in this way, the axioms directly characterize **P** over its entire field of application: that is both over the field of events – where it can be given the name of *probability* – and over the field of random quantities – where it is called, more generally, *prevision* (or price, if we are dealing with practical situations).

This is a great advantage, not only from a formal point of view but also because of the elegant simplification it provides. One avoids not merely the tiresome complication of having to consider two separate cases but also a whole series of difficulties that stem from the fact that such complications are misleading as well as annoying. In the first place, one encounters a tiresome complication if one wishes to formulate the axioms in such a way as to deal only with events, excluding random quantities. A further complication then arises when one attempts to put right this exclusion and define prevision, taking into account that it has already been defined in the particular case of events, where it is called probability.

The obvious way, and the only possible way, of dealing with the exclusion would simply be to remove it – even though not straightforwardly – by means of some device that puts us back on the straight and narrow. It seems, however, that the first, unhappy step obliges us to continue with it in making the second step. In wishing to consider as a *definition* of prevision some relation connecting it with probability, one is led into an extremely unnatural position. In other words, one makes it appear as though the elementary notion of prevision presupposes a knowledge of something much more complicated and delicate; that is, the probability *distribution* itself. Because it is unnatural, the situation is also dangerous, in the sense that it leads one to think that the definition to be made in this way, *ex novo*, allows a certain element of arbitrariness. In other words, that it requires, or permits, a choice of conventions, which are inspired by considerations of convenience.

In mathematical terms, expressed abstractly, all that we have said reduces to expressing a preference for, and then adopting, the first of the two paths open to us (which we indicate here by quoting the opening sentences of a more detailed description that can be found in Bodiou,[3] p. 5):

i) emphasis on linear functional (Riesz, Bourbaki, L. Schwartz);
ii) emphasis on measure (Borel, Lebesgue, Carathéodory, Fréchet, Kolmogorov).

2 If one limits oneself to X_h having a finite number of possible values; from here one could proceed to the general case (with bounded X_h) by means of approximations from above and below.
3 Georges Bodiou, *Théorie dialectique des probabilitiés englobant leur calcul classique et quantique,* Gauthiers-Villars, Paris (1964).

The main thing, however, is not the conclusion we reach – that is the choice itself – but rather the *reasons* lying behind this choice. It is not a question of saying which mathematical formulation has the greatest merit from a mathematical point of view, but rather of saying which provides a means of interpreting most directly those things which are most directly significant, most directly important, and, above all, most directly *observable* (in a conceptual sense).

Our attitude towards the difference between the two approaches to the definition of $\mathbf{P}(X)$ (the prevision of X, usually denoted by $\mathbf{E}(X)$ = the mathematical expectation of X) can be clarified by means of an analogy (which is, in fact, exact, apart from the change in terminology). Given a solid body C, one can define its 'barycentre', $B(C)$, say, and also give an operational method of determining it, without formulae; but it will not normally be possible (nor will it be important) to discover the mass distribution of C. In particular, the notion of 'density' at a point is simply a convention, defined by a limit process which, given the structure of matter (molecules, atoms, particles), cannot, strictly speaking, make any sense. However, it can be said that if we assume the density ρ to be known, as a function of the point P, we are then able to say that the mass of the body, $m(C)$, and its barycentre, $B(C)$, must be given by:

$$m(C) = \int_C \rho(P)\,\mathrm{d}S, \qquad B(C) = \frac{1}{m(C)}\int_C P.\rho(P)\,\mathrm{d}S.$$

To summarize: the difference we referred to consists of choosing between those definitions that are direct and intuitive, and those expressed in formulae as in the example above. (Note that in the latter case we require a passage to the limit in order to define density and then, to go back to the body itself, we have to do away with the density by integrating it. If there is any arbitrariness in the definition of the integral to be used, there is always the risk that some error is introduced.)

I find this undesirable habit of making simple things complicated to be very widespread at the present time (it is as if people go looking for trouble – and often they find it). I mention this not because I see it as my business to concern myself with it outside the confines of my own subject, but merely to point out that my noticing it and attempting to remedy it in the field of probability theory does not mean that I only see it as having taken root there. It happens more or less everywhere.

2 Events (true, false, and …)

By definition, an event must either be *true* or *false* (see Chapter 2, 2.3.4). It can be *uncertain* (for us, for the time being) only if, and insofar as, we do not possess the information required for establishing its truth or falsity. The same holds for any random entity; in particular, for random quantities. A 'random' quantity X is a quantity which has a well-determined value x; it could be, however, that we are not aware of what this value is (and it is because of this absence of information that it is, for us, for the time being, uncertain and, hence, random). We can, in fact, limit our discussion to the case of events, because any information concerning X is simply information concerning some event of the form $X \in I$ (where I is any set).

But what does it mean to say that an event is either true or false? Two extreme interpretations would consist in making reference to an 'objective truth' or to 'immediate verifiability'. The latter is unobjectionable but is extremely restrictive: it only holds in situations like that of a quiz where the answers can be found by turning to the next page. Even in this case, however, there are a number of implicit assumptions! We have to exclude the possibility of confusion or bewilderment such as would arise, for example, if every time one turned to the answers one found them different from when one last looked; or found them to be different according to whether one read them with the left eye or the right eye; and so on. Everyone will no doubt agree that these kinds of assumptions are ridiculous but it should be noted that there is no logical reason for regarding them as such. One does so because they conflict with certain 'regularities' that 'objective reality' has accustomed us to. (Dually, from a solipsistic point of view, they conflict with certain 'regularities' that have guided us in our construction of our idea of 'objective reality' – in the image of what appears to us in our maybe-real-world-maybe-dream-world.)

Should we let ourselves be guided by the objective interpretation, the first of the two extremes we mentioned above? Up to a certain point this is inevitable (otherwise we would be forever in the grip of a 'ridiculous' scepticism, as in the examples above). It is necessary, therefore, to be constantly on the alert, with a critical attitude, remembering that many statements that appeared to a 'naïve' objectivism to be undoubtedly meaningful had subsequently to be modified and revised in terms of 'operational' definitions in order for it to be possible to give them a meaning.[4]

But when is 'objectivism' not 'naïve'? Unfortunately, the answer is far from reassuring: 'it is so up until the point when the unexpected occurrence of the contradictions or drawbacks to which it gives rise actually take place'.[5] When this happens, one has to seek a remedy, and this consists in moving as far as we can in the opposite direction. In other words, we cease to think of the 'objective' fact of something being either true or false, but rather of the fact of whether or not we can obtain the information that for us determines whether it is true or not (or, at least, whether there is a possibility' – in some sense or other – of obtaining this information).

This would lead us to regard some events as worthy of the name (since it actually makes sense to ask whether they are true or false) and others as requiring elimination (in that they are bogus – events in appearance only, non-events). If the possibility of a clear-cut separation of the two kinds of events existed (or, at any rate, were assumed to exist – possibly with some appropriate, simplifying hypothesis), there would be no problem. Everything could then remain as before (including the definition of event),

4 At this point, in order to avoid confusion and misunderstandings, we should clarify the relationship between subjectivism in the field of probability and subjectivism in relation to knowledge in general. It is sometimes said that 'yes, of course probability is subjective … because *everything* is subjective'. Put this way, however, the statement is not in accordance with the subjectivistic conception of probability and is, in fact, at odds with it. The fundamental point of the subjectivistic conception is that the notion of probability does not refer to something which is a property of the 'outside world' (and it does not matter whether the latter is regarded as an 'objective reality' or as a 'mental construct'). A solipsist, who considered all of so-called 'reality' to be 'subjective', in order to be self-consistent, and to correctly interpret the subjectivistic concept of probability, would perhaps be right in saying, instead, that probability is *objective*. It is objective in the sense that it expresses an autonomous judgment and not something which is bound by 'external' circumstances to be interpreted in the sense of 'as if' (Veihinger's 'als ob').

5 It is almost the same as saying that every individual must be regarded as immortal until he eventually happens to die.

with an additional warning that one should make sure that one really is dealing with events (i.e. with events that make sense – those that are verifiable).

Instead, it seems to be necessary to retain more flexibility. More specifically (although this suggestion might appear to be an unhappy expedient), it will be convenient to use the term 'event' quite freely, without any *a priori* selection and exclusion. In this way, the selectivity can be brought in later, case by case, taking into account the different requirements (sometimes clear-cut, more often vague) that arise in connection with 'verifiability', and interpreting them in the light of appropriate (albeit to some extent arbitrary) schematizations.

It might be claimed that by adopting this approach we are begging the question, since the exclusion of that which *must* be excluded (because it is meaningless) becomes mixed up with the exclusion of that which *can* be excluded should we happen not to be interested in it. It is a fact, however, that our analysis (whether completely satisfying or not) does reveal the case of absolute unverifiability to be a limit-case of something more gradual (and, in a certain sense, 'economic'), involving different degrees of difficulty (of various kinds) in verifying whether an event is true or false. Nothing precludes one from evaluating this degree of difficulty in the light of the meaning and importance that such a verification would produce in practical terms.

One great advantage of proceeding in this way (and one which seems to me indispensable) is that our initial scheme remains the same: it includes all those things that we would have called events prior to embarking on these critical considerations, and permits us to carry out all the usual operations on them. When it comes to introducing a restriction (in a way which corresponds – within the given framework – to certain well-defined reasons), it will be sufficient to specify the subclass of events that one wishes to take into consideration (or to regard as making sense, or as verifiable, or whatever), and which other events one wishes to discard (regarding them as bogus, non-events, or as events whose meaning is unclear, or of little interest, or whatever). It may happen that in some circumstances certain given operations applied to verifiable events lead to verifiable events but that in other circumstances they do not. There are various possibilities of this kind and it is simply a question of noting what actually happens, rather than a theoretical question to be posed in abstract form as being an inherent feature of the concept of verifiability.

We have spoken thus far as if it were merely a question of distinguishing between genuine (i.e. verifiable) events, for which there are just two values, True and False, and bogus events, which are either not events at all, or are 'meaningless', or are 'intrinsically indeterminate'. There are cases, however, in which one discusses events for which there are three possibilities: True, False and Indeterminate (or Meaningless). This situation occurs above all in quantum mechanics in connection with the problem of complementarity and, hence, of indeterminism. Having three possibilities could give rise to a three-valued logic (as, for example, in Reichenbach, 1942).

In considering (in Chapter 4) *conditional events* of the form $E|H$, we were, in fact, dealing with logical entities that could take on three values: True ($1 = 1|1$; i.e. both H and E true); False ($0 = 0|1$; i.e. H true and E false); and Void ($\varnothing = 1|0 = 0|0$; i.e. H false and hence the truth or falsity of E irrelevant). This is precisely the way in which the 'three truth values' of the above-mentioned three-valued logic are formed (with H = 'an *observation* made in order to verify whether E occurs or not'; and only after this does it make any sense to ask whether E is true or false). In actual fact, this reduction of the problem to the simple and familiar set-up of conditional events does seem to provide an adequate solution; moreover, it is especially satisfying in that it avoids any formulation which might appear to contain the germ of metaphysical infection.

There is no problem if only one considers the meaningful components making up the conditional event $E|H$ to be not H and E, but H and EH (see Chapter 4, 4.4.1). So far as the event E in $E|H$ is concerned, it is immaterial whether we take it to be E, or EH (the minimum possible), or $EH + \tilde{H}$ (the maximum possible), or any intermediate event $E = EH + A$, with $A \subset \tilde{H}$. To ask whether E is meaningful (and if so whether it is true or false), when H is assumed false does not make sense, when considered in relation to $E|H$. In this context, one would be considering the question of whether or not the possible residual part of the sentence made sense, or was true or false. If it were meaningful at all, this would represent the irrelevant A, which is outside the field of interest. One could, however, investigate whether, for *other reasons*, the E in the formulation adopted – that is its residual part A which is irrelevant for $E|H$ – should be considered as meaningful and having interest outside of the hypothesis H. This is a separate problem, which concerns the event A as such, and does not have anything to do with the conditional event $E|H$, into which A enters only by the back door (like b in $5a - 0b + 2c$), or with its three logical values (for which, whatever A might be, there corresponds to $\tilde{H}$ – and hence to A and $\tilde{H} - A$ – always and only the same value; $\varnothing$ = Empty).

Let us now turn to a consideration of the mathematical representation of the field of events. We shall mention several variants, their appropriateness depending on the situation and on what is required. When it seems useful to do so, we shall also mention other possibilities that do not fit into our general framework of ideas. We have already provided a great deal of discussion in the text (Chapters 1–12) concerning the reasons for conflicting views on this subject; a few brief comments about certain peripheral topics should therefore suffice here. We shall try to give accurate accounts of those formulations that are not acceptable as such in terms of the approach adopted in the present work, and to bring out their worthwhile features, indicating how they might be applied in particular situations or as special cases.

3 Events in an Unrestricted Field

The basic set-up with which we shall begin is that already described in a summary fashion in Chapter 2, and which we have adhered to ever since, despite the occasional reservation. It serves our purpose in two ways: firstly, it is useful as it stands, in that it provides for a suitable representation and interpretation of the case that we shall regard as the most general, and, apparently, the simplest; secondly, with appropriate modifications, it provides a means of obtaining schemes for representing the other cases of interest (and some aspects of these might, in fact, appear simpler than the first case).

The simplicity of this first case lies precisely in the fact that no restrictions of any kind are imposed when it comes to forming the events: the latter could always be thought of as *arbitrary subsets* of a set of 'elementary possible cases' of a partition admitting indefinite refinements. In formal terms, we could express this more precisely as follows: 'at any given moment', the field $\mathscr{E}$ of events under consideration corresponds to the entire collection of subsets (or subdivisions) of the set (or *partition*) $\mathcal{Q}$ of the 'elementary possible cases', Q, which, 'at that moment', one wishes to single out. The partition $\mathcal{Q}$ must be considered as having no structure whatsoever and, moreover, it must be considered as 'provisional', 'not once-and-for-all' (and this is what the references to a 'given moment' etc. are intended to convey). This implies that we can only consider as meaningful those notions and properties which are, in a certain sense, invariant with respect to 'refinements' of $\mathcal{Q}$. In more precise forms the latter are represented by subsets $\mathcal{Q}'$, for which each set of $\mathcal{Q}$ consisting of a single point Q is replaced by a set containing many 'points' Q' (in general, there could be an infinite number). To summarize: we can provisionally identify the events E of $\mathscr{E}$ with the subsets $\mathfrak{P}(\mathcal{Q})$ of $\mathcal{Q}$, and also with the corresponding subsets $\mathfrak{P}(\mathrm{Q}')$ of Q'; – that is without taking too seriously the temporary interpretation of the $Q \in \mathcal{Q}$ as 'points'.

We shall give an example straightaway, in order to clarify this. Let E be the event $X^2 + Y^2 \leqslant a^2$ (where X and Y are random quantities). If $\mathcal{Q}$ is the (x, y)-plane, the event E corresponds to the disc of 'points' (x, y) with $x^2 + y^2 \leqslant a^2$; whereas if $\mathcal{Q}'$ is the three-dimensional space of points (x, y, z) (or four-dimensional, (x, y, z, t) etc.) the event corresponds to the cylinder of points (x, y, z) (or x, y, z, t) etc.) such that $x^2 + y^2 \leqslant a^2$. By considering not only the random quantities X and Y, but also others like Z and T, etc., we change the field $\mathcal{Q}$ and the notion of point (to each point (x_0, y_0) there corresponds the infinity of points on the line (x_0, y_0, z), or on the plane (x_0, y_0, z, t), etc.). The set to which E corresponds – or, conventionally, with which it is considered identified – changes, but this is an irrelevant contingency, arising from the form of representation; what does not change is the meaning of the proposition itself, which is completely contained in the inequality $X^2 + Y^2 \leqslant a^2$.

All this could have been expressed in a better way had we eliminated completely the notion of point, but it appears to be more instructive to put it forward and then to present the arguments against it. In this way, we underline the contrast between the more usual formulations on the one hand, and the refusal to accept 'closure' – as is otherwise inevitable – on the other.[6] On the other hand, it turns out to be useful to accept the 'points' as indicating the limit of subdivision beyond which it is not necessary to proceed (at a given 'moment', i.e. with respect to the problems under consideration). There is just one condition: that we always bear in mind that this is only useful

6 To approach the formulation of a theory by starting off with a preassigned, rigid and 'closed' scheme seems to me a tiresome and cumbersome procedure, wherever it is followed. (It is true that it serves to guarantee one against antinomies and suchlike, but this is not a good reason for always having recourse to it; in the same way as it is not necessary to shut oneself inside a tank in order to journey through a peaceful and friendly country.)

In connection with the use of 'points', and their abandonment in geometrical representations, we refer the reader back to our remarks in Chapter 2, 2.4.3 (especially to the quotations of von Neumann and Ulam).

Further discussion, closely relevant to this point, and (insofar as the present topic is concerned) upholding precisely the same position, can be found in Bodiou (1964, p. 3). Abstracting from the space of points, and describing the events directly as the elements of a Boolean algebra, or a Boolean lattice, he observes that 'apart from the formal simplification thus obtained, the new axiomatization is more directly interpretable in terms of the logic of the attributes.... If one assumes that each element of the Boolean algebra is a union of "atoms", one proves equivalence to the Kolmogorov axioms; *the emphasis, however, is more directly on the essential feature – the global lattice, and not the set of its atoms*'. (We should add – although it is not necessary at this point – that Bodiou does, however, retain the usual conditions, which admit σ-additivity.)

insofar as it helps to 'fix ideas' at the time in question. If one were to attribute to it some absolute meaning, it would lead to a confusion of the ideas, and to a tangle of misunderstandings.[7]

The field $\mathscr{Q}$ (or, more generally, a field $\mathscr{S}$ of which $\mathscr{Q}$ is just a part) is often obtained by starting from some given set – which we shall call a basis $\mathscr{B}$ – of events E_h, or, more generally, of random quantities X_h (this was the case in the previous example, where we started with X and Y, and then considered adding in Z and T). To think of the field Q (and therefore of the field $\mathscr{E} = \mathfrak{P}(\mathscr{Q})$) as having been generated from a basis B is completely irrelevant; it is convenient, however, to refer to this case in order to take up the theoretical discussion again, and to develop it in a more expressive manner.

In Chapter 2, and also in the example above, we have already seen how, given n random quantities X_h, the whole picture could be summed up by considering a single 'random point' Q in the Cartesian space $\mathscr{S} = S_n$ (with coordinate system x_h), where Q is the point defined by $x_h = X_h (h = 1, 2,\ldots, n)$. Not all the points of $\mathscr{S}$ are, in general, *possible*, but only those of a subset Q (obtained by eliminating the cases $\mathscr{S} - \mathscr{Q}$ which, on the basis of the data of the problem, turn out to be impossible). In particular, if the X_h are events, $X_h = E_h$, $\mathscr{S}$ reduces to the set of the 2^n vertices of the hypercube (with coordinates 0 or 1), since we can only have $x_h = 0$ or $x_h = 1$. In this case, $\mathscr{Q}$ is the subset of possible vertices; in other words, the constituents (Chapter 2, 2.7.1). In the general case, nothing really changes, except that the E_h, or the X_h, may be infinite in number; the indices h will then run through some infinite set H (not necessarily countable), and even if we write the more familiar $h = 1, 2, 3,\ldots$, or simply say 'all the E_h (or X_h)', we shall mean $h \in H$.

In this way, the preceding (Cartesian) representation will hold without any alteration, except that the number of dimensions (of axes, of coordinates) is infinite,[8] and $\mathscr{S}$ will be S_H (the cartesian space with an infinite number of coordinates, x_h, $h \in H$). In the case of events (i.e. if all the X_h reduce to events E_h) the vertices of the hypercube (in infinite dimensions) are characterized by indicating for which h we have $E_h = x_h = 1$ (for the others, $E_h = x_h = 0$). In other words, they correspond to subsets of $\mathscr{B}$ ($\mathscr{S} = \mathfrak{P}(\mathscr{B})$) or, equivalently, to the functions $f(\cdot)$, elements of $\mathscr{S} = 2^{\mathscr{B}}$, which to some subsets of $\mathscr{B}$ assign the value 1, and to the others 0. One easily recognizes the identical form of procedure to that which led to constituents in the case of finite n; it is a question of stating that out of the events of the basis $\mathscr{B}$, a certain subset are true, and the others false. Of course, some of the products will, in general, be impossible; that is demonstrably false on the basis of the data. We shall need to remove these from $\mathscr{S}$ in order to obtain $\mathscr{Q}$. If we wish, we can always reduce the

7 The most serious such misunderstanding likely to arise is the idea that a conditional event $E|H$ has some special significance when H is 'atomic': in other words, when H corresponds to a 'point' in some given representation (although this would obviously be incomplete, since both EH and $\tilde{\mathscr{E}}\mathscr{H}$ must make sense, and $H = EH + \tilde{\mathscr{E}}\mathscr{H}$). In this way, one would be led to think that $\mathbf{P}(E|H)$ has an absolute meaning, unchanged even if some further information can be added to that expressed by H (here, $\mathbf{P}(E|H)$ is the probability of E, 'knowing all the circumstances that can influence E – and, one might add, determined up to the present moment – as expressed by H').

Considering the early stage we are at in our present attempt at a systemization, the above brief comments may seem premature. However, it is perhaps useful to have some idea of the arguments we shall have to consider, even though we shall only come across them later, while developing this treatment.

8 Note (although this is not really important here) that the concept of the number of dimensions – when this is infinite – could be understood in a different way as the number of nonzero linearly independent elements (and this is actually intrinsically more meaningful): this notion no longer coincides – as in the case of finite n – with the 'number of coordinates'.

case of random quantities to that of events: it suffices to substitute for each X_h the events $E_{h,x} = (X_h = x)$, for all the values x possible for X_h. Calling $\mathscr{B}'$ the modified basis which arises from this substitution for all the X_h, we can always write S = $\mathfrak{P}(\mathscr{B}')$, or $\mathscr{S} = 2^{\mathscr{B}'}$.

If, having constructed $\mathscr{S}$ in this way (using one variant or another), we preserve, even if implicitly, through the x_h, the record of how $\mathscr{S}$ was generated from the basis $\mathscr{B}$, a linear space structure (or that of a subspace) remains as a trace of this in $\mathscr{S}$ (and hence in $\mathscr{Q}$). On the other hand, we might actually be dealing with a problem of geometrical probability (even of geometry, in the sense of ordinary, physical space) and hence we inevitably have the geometric structure (one could think, for example of $\mathscr{S} = \mathscr{Q}$ = surface of the earth, Q = point at which a lost – or stolen – object is located). It does not matter. In saying that Q has to be considered as having no structure, one is not saying that a structure might not be seen to exist if we looked at it *from a different standpoint* (and one does not wish to rule out the possibility of taking this into account if it should appear at a later stage to suit our purpose to do so). We simply mean that *for the time being and for our present purpose* we must ignore it.

If we do not choose to ignore the way in which $\mathscr{S}$ has been derived from the basis $\mathscr{B}$, the possibility arises that we could single out certain events as being somewhat *special*: for example, belonging to the basis, or logically expressible in terms of a finite or countable number of basis elements. Similar distinctions could be drawn among random quantities: those belonging to the basis, or functions of basis elements (the functions being linear, or continuous, or whatever, and involving any particular number of basis elements), and so on. This is why on the real line, starting from the intervals (as basis), one is able to maintain distinctions between sets that are sums of a finite, or countable, number of intervals, or obtainable from intervals by at most a countable number of logical operations, and those which are not. We mention this familiar example merely in order to point out that the introduction of ideas of this kind, and the consideration of such distinctions, is not admissible, and, in fact, must be explicitly excluded, since we wish to regard $\mathscr{Q}$ as having no structure whatsoever (at least for the time being).

There remains just one distinction – a nuisance as far as we are concerned – which, at the present time, would continue to make sense, even if we consider $\mathscr{Q}$ as having no structure. It is that based on the *number* of elements ('points') of the given E and of its complement $\tilde{E}$. More precisely, if the cardinality[9] of the whole set $\mathscr{Q} = E \cup \tilde{E}$ is M (infinite), then either E and $\tilde{E}$ both have the same cardinal, M, or one of them has cardinal M, and the cardinal of the other is smaller (either 0, 1, 2,..., n,..., or $\mathfrak{n}$ – countable infinity – if $M > \mathfrak{n}$, or some other cardinal N, where $\mathfrak{n} < N < M$). The introduction of the convention of 'never' regarding the subdivision as a 'final' one is another reason for ignoring the structural distinctions. In speaking of the *basis*, we can express this by saying that we must always be aware that other events (or random quantities) can be added at will. In this way, the distinction based on the *number* of 'points' also becomes

9 We denote by $\mathfrak{n}$ and $\mathfrak{c}$ (Gothic *n* and *c*) the cardinals of the *integers* (the smallest infinite cardinal) and of the *reals* (the continuum), respectively. If A and B are (disjoint) sets with cardinals M and N, respectively, then the cardinals of $A \cup B$ (union), $A \times B$ (Cartesian product of ordered pairs (a, b)) and A^B (the set of functions from B to A), are given by $M + N$, MN and M^N, respectively. If M and N are infinite, and $M > N$, we have $M + N = MN = M^N = M$ (and also $M + M = MM = M$): however, $2^M > M$ (and, *a fortiori*, $N^M > M$), and, in particular, $2^{\mathfrak{n}} = \mathfrak{c} > \mathfrak{n}$ (there are as many subsets of the integers as there are points on the real line).

We shall be making use of these properties in what follows, hence the reason for our recalling them here.

meaningless (and this applies, in particular, to the distinction between 'atomic' events – corresponding to single points – and others). There remains the one structural distinction that we must, of course, retain as a meaningful one: that between the *impossible* event ($E \equiv 0$, corresponding to the empty set), the *certain* event ($E \equiv 1$, corresponding to the entire set), and all the others (the *possible* events, which are structurally indistinguishable among themselves).

If one actually wishes to continue along these lines, by introducing new elements into the basis, or – if one prefers to put it this way – passing from $\mathscr{Q}$ to $\mathscr{Q}' = \mathscr{Q} \times \mathscr{Q}^*$ (the cartesian product with any suitable $\mathscr{Q}^*$) until each 'elementary case' Q of $\mathscr{Q}$ is subdivided into M elements, all the E (thought of as sets of $\mathscr{Q}'$), apart from the empty set, will have cardinality M (because the cardinality will be both $\geqslant M$ and $\leqslant M \cdot M = M$). In this way, only the empty set and its complement will be distinguishable from the others. It would, however, be cumbersome to actually reduce oneself to such a $\mathscr{Q}'$. We shall content ourselves, therefore, with the fact of having mentioned the possibility of this equalizing of the cardinality for all the events of interest, without insisting on it being done, or taking it into any further account. It is sufficient to state that we ignore as being irrelevant any distinctions made on the basis of considering cardinality. We shall not mention this again. In fact, without going into all the details, all this could have been regarded as implicit in the assertion that we were not going to acknowledge any distinction between the subdivisions of a partition Q, and the (corresponding) subdivisions of a finer partition $\mathscr{Q}'$.

Instead, we must go back to the problem of the nuisance structures introduced by the presence of the basis $\mathscr{B}$; the structures that we had decided to *ignore*. Rather than ignoring them, we can *make use* of them, by removing, in a different way, the drawback they had of inducing a special status for some events, or random quantities, in comparison with the others. Instead of prescribing that the basis be ignored (and let us suppose for the moment that we have a basis of events, $E_h \in \mathscr{B}$, $h \in H$), we can achieve the desired result by enriching the basis itself so that it includes all the $E \in \mathscr{E}$. It then follows that membership of the original basis $\mathscr{B}$ is no longer relevant.[10] In the case of a basis of random quantities ($X_{h \in} \mathscr{B}$, where $h \in H$), a thorough application of this same procedure takes us even further. More precisely, it will be a question of adding to the X_h of the basis $\mathscr{B}$ all the random quantities expressible as functions of them (*any* functions); that is every function X_k of the points $Q \in \mathscr{Q}$, $X_k = f_k(Q)$, where the $f_k(\cdot)$, $k \in K$, put every Q of $\mathscr{Q}$ into correspondence with a real number. (It goes without saying that we do not impose any restrictions like continuity, etc., because we have already said, and repeated, that we do not consider $\mathscr{Q}$ as having any structure, and so such restrictions do not even make sense.[11]) The field $\mathscr{S}$ will then be the Cartesian space – let us denote it by S_K – with an infinite number of coordinates $x_k (k \in K)$, where K is the set of the indices k that label the functions $f_k(\cdot)$ forming the field $\mathscr{S} = \mathfrak{c}^{\mathscr{Q}}$. Essentially, K is $\mathscr{S}$ itself (and it is only for

10 This is rather like everyone at birth receiving the title of 'Your Excellency' in order to achieve its downgrading (something which the abolition of the title would not, since there would always be a handful of people who would retain it).

11 The distinctions which clearly do make sense are those concerning the 'possible' values of the X_k (i.e. the range of $f_k(Q)$); the case of *bounded* X_k is particularly important (as we have often seen already in this work). Here we are not directly interested in this aspect, because it does not depend on things concerning the field $\mathscr{Q}$ (and related notions).

notational convenience that we introduce an index k, $k \in K$, in order to distinguish the functions $f_k(\cdot)$; it would be equivalent to refer to the functions f, $f \in \mathscr{S}$). One may observe that K (or, equivalently, S) has cardinal $\mathfrak{c}^M$, where M is the cardinal of $\mathscr{Q}$, and that K contains H, $H \subset K$, given that among the complete collection of $X_k = f_k(Q)$ there exist, in particular, the X_h which form the basis $\mathscr{B}$.

The 'waste' in the number of dimensions (in passing from H to K) is clearly considerable and might seem rather absurd. On the one hand, however, the fact that this enormous waste is more or less a disaster is irrelevant in practice, since it is never necessary, nor would it be possible, to take account of the infinite dimensions one at a time. On the other hand, such an extension brings with it something that is very much to our advantage (formally, for the time being, but of substance later, when we introduce into this framework the notions of probability and prevision). This advantage lies in the following: that, *by this means*, we could also *retain, and consider as valid, the linear structure* thus introduced into $\mathscr{S}$ and which makes it into a linear ambit $\mathscr{A}$, since – by the principle of 'everyone a nobleman' – it *no longer gives rise to any discrimination among the various random quantities and, in particular, among the events.*[12]

The extension made for $\mathscr{S}$ does not modify, in any basic respect, the set, or field, Q of possible points Q. Those that are (provisionally) considered as 'elementary possible cases' remain the same, but the 'points' representing them are dispersed and spread in the enlarged field $\mathscr{S}$ to a much greater extent. A basic intuition can be obtained from recalling the example of the parabola, $y = x^2$, given in Chapter 2, 2.8.7; however, this only concerns a single random quantity. In the general case (and also in the case of the example above, where one considers all the $Y = f(X)$, even restricting oneself to $Y = X^n$), it turns out that all the points Q (in the field $\mathscr{S}$ extended to a linear ambit A) are *linearly independent*. In other words, if $Q_1, Q_2, \ldots, Q_n$ are *possible* points – belonging to Q – then, in all the S_{n-1} which they determine, there is no other possible point (i.e. the intersection of $\mathscr{Q}$ with such S_{n-1} reduces to these n points, and in general consists of at most n points).

In order to fix ideas, let us verify this first of all for the simplest example, which we mentioned above. The field $\mathscr{S}$ is the Cartesian space with an infinite number of coordinates, x_h ($h = 1, 2, 3, \ldots$), on which are represented the values of the random quantities $X_h = T^h$ (where, for greater clarity, we denote by T the random quantity with which we begin; i.e. X_1, represented on the x_1-axis). The field $\mathscr{Q}$ is the 'line'[13] with parametric equations $x_1 = t, x_2 = t^2, x_3 = t^3, \ldots, x_h = t^h, \ldots$, if the random quantity T admits all the reals

12 In Bodiou (1964, *op. cit.*) we find further discussion of these topics, again in agreement with our views, and all the more interesting since his discussion is not inspired by abstract, conceptual questions like those we have raised here, but by problems in quantum mechanics. Arguing in favour of referring to a 'dialectic lattice' (like the one he proposes) rather than to a particular special form of Hilbert space, he makes the following remark (p. 103): 'The unwarranted special status conferred on the coordinates of the particle by this particularization obscures the general character of the notions, and gives rise to pseudo-problems (…) The coordinates are random quantities just like all the others, no matter how important they might seem.'

13 We use the word 'line' for convenience, it being a set of points depending on $t(-\infty < t < +\infty)$. For our purposes, it does not matter whether this term is really appropriate for some other aspects of the problem. (One thinks, for example, of the 'peculiar fact' that on the segments where $|t| > 1$, the points of the 'line', with the usual metric, all have infinite distance from one another. However, it would be sufficient to consider the modified line, $x_h = t^h/h!$, or, equivalently, to use the metric $[\sum_h (x_h / h!)^2]^{\frac{1}{2}}$, in order to overcome these difficulties. We mention all this merely for the sake of curiosity.)

(from $-\infty$ to $+\infty$) as possible values; otherwise, it is the subset of the points of the 'line' corresponding to the values $t \in I$ of the parameter t belonging to the set I of possible values for T (one should constantly refer back to the example in Chapter 2, 2.8.7).

To establish that linear independence exists between the points of $\mathscr{Q}$, it is sufficient to recall, for example, the fact that the Vandermonde determinant is nonzero. Given any n points of $\mathscr{Q}$ – $Q_1, Q_2,\ldots, Q_n$, say, corresponding to $t = t_1, t_2,\ldots, t_n$ – if they were linearly dependent (i.e. if they belonged to an S_m, $m < n - 1$) then, *a fortiori*, their projections onto an $(n - 1)$-dimensional subspace would also be linearly dependent. One could, for example, take the projections onto the subspace obtained by considering only the first $n - 1$ coordinates $x_1, x_2,\ldots, x_{n-1}$ (setting $x_h = 0$, $h \geqslant n$). This would imply the vanishing of the Vandermonde determinant ($a_{rs} = t_r^{s-1}; r,s = 1,2,\ldots,r$), which is impossible for distinct values of t. It might be observed that the same proof also holds for all other infinite projections (but one is enough to establish the conclusion).

The proof of linear independence in the case of a general linear ambit A is even simpler: the preceding case is useful only in that it deals with a situation that is, in a certain sense, more immediate, because no discontinuities are involved (these arise in the consideration of events, with 0 and 1 as the only possible values). Let $Q_1, Q_2,\ldots, Q_n$ be once again points of $\mathscr{Q}$, and let Q_0 be a point which is linearly dependent on them:

$$Q_0 = a_1Q_1 + a_2Q_2 + \ldots + a_nQ_n$$

with $\Sigma a_h = 1$.[14] Let us divide the n points Q_h into two groups, labelling them with one or two dashes, respectively (i.e. writing Q'_h or Q''_h to indicate whether Q_h is in the first or second group). The only condition is that the sums a' and a'' of the weights a'_h and a''_h are neither 0 nor 1 (this can always be arranged, except in the case in which a single a_h is equal to 1, and all the others are zero; i.e. the case in which Q_0 coincides with one of the given points Q_h: we shall obviously exclude this case). We denote by E' and E'' the logical sums of the Q' and the Q'', and we let E be any event $E' \subset E \subset \tilde{E}''$ (in simple terms, we put the Q' in E, the Q'' in $\tilde{E}$, and we divide up all the other points of Q arbitrarily between E and $\tilde{E}$). For any such event E, we can say that it has the value 1 over all the points Q', and the value 0 over all the Q''; consequently, it has value a' on Q_0 (where $0 \neq a' \neq 1$). It follows that Q_0 is not a point of $\mathscr{Q}$ (and therefore not 'possible') because it does not attribute to E one of the two values 0 or 1.

It might well seem absurd to 'invent' – not without some effort – a field S, or a linear ambit $\mathscr{A}$, constituted almost entirely of points that satisfy ridiculous conditions (like making an event – whose values can only be 0 or 1 – assume values such as log 2, or π; or making X take on the value 1, and X^2 take on 2 or 0). It may be, however, that this makes sense in terms of probability and prevision (we might well have $\mathbf{P}(E) = \log 2$ or $\mathbf{P}(X) = 1$ with $\mathbf{P}(X^2) = 2$; this would happen, for example, if X took on the values 0 and 2, each with probability $\frac{1}{2}$), or in terms of linear combinations of previsions (see the footnote to Chapter 3, 3.7.2). Let us take advantage of this glimpse into the future in order to simplify our construction somewhat. Suppose that we assume (rather unjustifiably, because it runs a little ahead of the axiomatic treatment) that points satisfying

14 This notation really implies that Q_0 is the barycentre of the Q_h with masses a_h; i.e. the point whose barycentric coordinates, with reference to the basis-points Q_h, are a_h. In Cartesian coordinates, this implies that, for any x, the values $x^{(i)}$ which it assumes at Q_i ($i = 0, 1, 2,\ldots, n$) satisfy a similar relationship, $x^{(0)} = a_1x^{(1)} + a_2x^{(2)} + \ldots + a_nx^{(n)}$.

linearly contradictory conditions are of no use (for example, $2X - Y = 3$, $X + 4Y = 1$, $3X - 6Y = 4$; we see that $3X - 6Y = 2(2X - Y) - (X + 4Y) = 2.3 - 1 = 6 - 1 = 5$). We can then restrict the linear ambits $\mathscr{A}$ to that part (of the larger construction) where things go through for linear relations.

As an alternative to this *a posteriori* exclusion of the superfluous part, one could simply avoid constructing it in the first place. Thinking in terms of a transfinite mode of construction, introducing one after the other the functions $X_k = f_k(Q)$, it is sufficient not to introduce new axes x_k whenever the X_k turns out to be a (finite) linear combination of those already considered. We shall make the convention that the (bogus) random quantity $X_0 \equiv 1$ be introduced as the first element (for the same reasons, and with the same effects, as in Chapter 2, 2.8.3) and we shall obtain the *linear space* $\mathscr{L}$ of the X, the dual of the linear ambit $\mathscr{A}$ (this follows in the same way as before, so we omit the details).

In order to illustrate all this in a more concrete fashion, it is convenient to refer to the construction we mentioned above. Some of the X_k (let us denote them by X_k^*) are represented on new coordinates x_k, whereas those linearly expressible in terms of a finite number of the preceding ones already find a coordinate available: $x = u_0 x_0 + u_{k_1} x_{k_1} + \ldots + u_{k_n} x_{k_n}$. This representation is one-to-one[15] and also holds for the preceding case (for $X_k^* : x_k = u_k x_k, u_k = 1$), which does not have to be regarded as special in any way. Each point Q, and similarly each point A of $\mathscr{A}$ (even those not possible), is characterized by the values $x_k(Q)$, respectively, $x_k(A)$, of its coordinates on the axes of the X_k^*. For every other X, the value will be given by the above finite linear combinations (of the coordinates x_{k_i}, calculated at Q or at A, respectively). These values could be written (just as in Chapter 2, 2.8.3) as $A(X)$ or $X(A)$, and interpreted as products of vectors (from $\mathscr{L}$ and from $\mathscr{A}$), with the sole difference that instead of linear combinations from among a finite number of elements ($n + 1$), we have finite linear combinations from among an infinite number of elements.

We shall call a halt here to our description of the formal set-up. The consideration of these topics – although necessary for a complete presentation of the scheme – has already led us so far into the probabilistic meaning that we cannot usefully say anything more without bringing in the latter explicitly.

4 Questions Concerning 'Possibility'

The most clear-cut distinction in the entire formulation is that between possible and impossible events (or certain ones, but these are merely complements of the impossible ones). But is this distinction really so clear-cut?

There would seem little room for doubt. Someone who argues that he does not know whether an event is possible or not is, in actual fact, already saying that for him it is possible (because he cannot exclude it, as would be necessary if it were impossible). In the same way, when one says that 'it is not known whether, in a census, the sex of a particular individual is known or not', one is already admitting that it is 'unknown'. As a matter of terminology, this is without doubt absolutely correct. When we try to apply

15 Were it not so, there would exist a linear relation among a finite number of elements X_k^*, and then the *last* one would have been mistaken for an X^*.

the principle to actual situations, however, and we start examining the exact nature of the dichotomy between what one *knows* and what one does *not know*, the dividing line seems much less absolute and various kinds of difficulties arise.

The most serious of these is the one concealed in the very logical mechanisms which exist to overcome it. We know that every consequence of something which is certain is itself certain; in other words, all that is implicit in those things for which we have explicit information must be considered as part of that information. It follows that the field of what is certain for someone – that is for which information is available – must be *closed* with respect to deduction; in other words, it must not leave outside anything which is deducible. But deducible in what sense? Through the mechanisms of logic, and this, as Galileo says in a celebrated passage, requires 'voyages of our mind, step by step, with time and with motion'; but for things to be as easily done as said, we should require 'the mind of the Almighty', voyaging 'with the speed of light'.

The extreme case, in an opposite sense, is that in which, for some individual, certain conclusions are completely beyond reach; either because his knowledge is insufficient, or because he is not capable of the necessary reasoning. Suppose that N is the number of paving stones forming some given rectangular pavement (for example, one that the individual recalls having seen, but for which he has only a vague recollection of the dimensions, and of the dimensions of the paving stones). The set of possible values for N should, at the very least, exclude the prime numbers. But what if he is not familiar with this notion? Or if he was familiar with it once upon a time, but finds that things he learnt at school do not come back to him sufficiently readily when he is faced with problems where such knowledge would be useful? In these circumstances, he could not even contemplate making such an exclusion. This extreme case is in any case the simplest; the set of things that are 'certain' always remains closed with respect to this deductive capability (even if in an incoherent fashion, since it does not coincide with the logical possibilities).

The situation is more awkward when 'closure' fails to hold and is replaced by something less rigid. The clear-cut distinction between those certainties that an individual can work out for himself and those which he cannot is lost, and instead we have certainties that are attainable with various degrees of difficulty, and perhaps not immediately. In terms of the above example, this is the situation faced by an individual who knows that he has to exclude the prime numbers, but who finds that for many integers it is not easy to see at a glance whether or not they are prime. Is it worthwhile making all the calculations required in order to find out? Or is it worthwhile checking through a table of prime numbers (first searching for such a table and then searching for the number in question)? It would probably not be worthwhile if the purpose were merely to exclude certain rare numbers from the set of those to be considered as 'possible'. It might be the case, however, that, by virtue of some additional information, we were in a position to determine N uniquely, since it would turn out to be the only possible value. (As an example, suppose we knew that the diagonal of the rectangle were a multiple of the side of the paving stones. It follows that $N = XY$, where the sum of the squares of X and Y is a square; if we had sufficiently close bounds, $x' \leqslant X \leqslant x''$, $y' \leqslant Y \leqslant y''$, or some other similar information, X, Y, N would turn out to be uniquely determined.) In such cases 'is it worthwhile'?

In order to answer this question, we need, of course, to know what the alternatives are, and what degree of interest they hold for the individual relative to his evaluation of the difficulty and the labour of finding out. Among the possibilities, we note the following: we are not interested in the problem, not even out of curiosity, and we have the option of not bothering ourselves with it; we are content with a crude estimation (for example, we may wish to buy n paving stones, choosing n such that we are almost certain of there being sufficient to enable us to form a pavement of the same size as the original, but with not more than 10% of the paving stones left over; $\mathbf{P}(N > n) = 1\%$, $\mathbf{P}(N < 0{\cdot}9n) \leqslant 5\%$, say); if, instead, there is a lot at stake, we can make precise our distribution of probability, $p_h = \mathbf{P}(N = h)$, and compare the costs, risks and advantages of buying this or that number of paving stones.

The aspect that concerns us most here is the difficulty of verifying just what it is that is implicit in our initial data. Sometimes this difficulty will be insurmountable (evaluate the billionth decimal place of π; check the Goldbach conjecture up to 10^{100} etc.; and it would be even worse were we to consider things of this kind involving an infinite number of digits or integers); at other times, the difficulty is simply one of the labour or cost involved. We shall encounter these matters again in connection with the verifiability of events, and we shall then go into them more deeply. At this juncture, it is sufficient to observe that, from a practical point of view, there is, in the last analysis, no difference between an experiment or investigation aimed at uncovering information about an unknown fact, and an attempt at deduction aimed at ascertaining that which, in theory, we should already know on the basis of the information in our possession. The difference is simply that if we abandon such attempts at deduction (because of more or less insurmountable difficulties), the field of that which we 'actually know' is incoherent, since it is not closed from a logical point of view.

We obtain maximum 'flexibility' if it is easy to demarcate the field of those cases 'possible at first sight'; the difficulty here, however, is that it takes a long time to check all cases. (To give a simple, though rather silly, example: all the integers n between 1 and 1000 excluding those for which 5 appears at least twice in the first six decimal places of the reciprocal.) To check through to the very end presents difficulties, but the same is true of stopping at some arbitrary point and, therefore, of starting off with any rigorously laid down distinction between those cases which are possible and those which are not.

Such a distinction becomes even weaker if we abandon the convenient hypothesis that the knowledge one starts with is precisely specified in the stated 'data of the problem'. If we trace back to the actual, empirical source of the knowledge, are we really able to draw a reasonable line between possible and impossible? If a person had seen our pavement he might very well say that he is 'certain' (but is it true?) that it could neither be less than $1 \times 2\,\mathrm{m}^2$, nor greater than $100 \times 1000\,\mathrm{m}^2$ (suppose he estimated it as about 4×7). But can such an absolute logical distinction be declared valid in one case and yet one not know at what point it ceases to be valid? Is the length of 3 m too small to be possible? Yes? And a length of 4 m? No? Then what about the limit being 3·42 or 3·423 m? Can it not be specified precisely?

It would perhaps be appropriate to only use the phrase 'absolutely certain' when referring to tautologies (and even then …?). Adopting this rule in our example, we could only say that the side may have any length between 0 and ∞. In fact, nothing prevents us, either for convenience or by convention, from restricting our consideration of the problem to a narrower and more 'reasonable' set of values (like those mentioned above, 1×2 and 100×1000, or to an even more restricted set). It would, however, be more prudent and responsible to substitute in place of 'it is certain that …', the phrase 'assuming that … (the risk of my being wrong appears to me completely negligible)'. Can we talk of an age limit beyond which it is certain that some given individual (or any person living at the present time) cannot survive? Or of a speed which cannot be exceeded (for example, in a track race, or swimming 1000 m, or by a bicycle, car, aeroplane etc.)? I would say that it is always inadvisable to express in terms of 'certainty' any everyday judgement of this kind (even on those occasions when I share the judgement, though with a less firm conviction).

So what conclusion do we come to? We could well repeat the statement that we made initially: that if we do not know whether something is possible or impossible, then, by definition, it is possible. Let us bear in mind, however, that everything is based on distinctions that are themselves uncertain and vague, and which we conventionally translate into terms of certainty only because of the logical formulation. On the other hand, even in the case of those in a census for whom 'it is not known whether or not the sex is known', it may very well be that the doubt does not have to be resolved straightforwardly by saying 'therefore it is unknown'. This is true, certainly, but if we are considering the possibility of further investigation, allowing – either definitely or potentially – the completion of the missing information, the nature of the problem changes.

In the mathematical formulation of any problem it is necessary to base oneself on some appropriate idealizations and simplifications. This is, however, a disadvantage; it is a distorting factor which one should always try to keep in check, and to approach circumspectly. It is unfortunate that the reverse often happens. One loses sight of the original nature of the problem, falls in love with the idealization, and then blames reality for not conforming to it.

In our case, it is certainly necessary that we base ourselves, in the initial formulation, on the distinction between possible and impossible; the distinction being considered as clear-cut as suits our purpose. We must be on our guard, however, not to become prisoners of this artificial rigidity (of absolute certainty) if, rather than helping us, it should come to trap us in an incomplete and distorted view of things. We shall return to these questions, and delve more deeply into them, both in relation to impossibility (as above) and with (new) reference to zero probability, and, above all, in relation to the acquisition of information for the purpose of an evaluation of probability. It is only then that we may come to have a more precise understanding of the questions we have had to illustrate here in terms of one particular aspect, without considering the complementary one.

5 Verifiability and the Time Factor

We can admit that an event *E*, suitably described, makes sense objectively; that is, is true or false independently of any possibility that we have of knowing the fact. To affirm or deny its truth as a general thesis would mean that one was being metaphysical, but it cannot be doubted that it is often convenient and almost unavoidable to think in this way (albeit with due caution).

However, what is important for our purposes is not the fact that *E* is objectively true or false (if one can speak of a 'fact' in this connection). What really matters is to establish the fact (to obtain the information, to verify) that *E* is true or *E* is false. Think in terms of a bet involving *E*: this is the most futile, but nonetheless the most expressive, example for demonstrating that what counts is only the knowledge of the outcome. Moreover, the case of a bet serves as a typical model of the general situation in which probabilities serve as guidelines for decision making under uncertainty and for applications within this framework.

Let us return now to the study of the field of possibility, which we began in Section 3 (without restrictions; and let us leave aside, for the time being, the reservations made in Section 4). Let us assume that in this field every event is – in an 'objective' sense – either true or false (and that every random quantity X has a precisely determined value, x); in other words, we assume that one may imagine as uniquely defined that point Q of $\mathscr{Q}$ which summarizes the truth or falsity of all the events of $\mathscr{E}$ and the exact values of all the random quantities of $\mathscr{L}$.

It might be thought that such a schematization is excessively theoretical and pretentious, even for the representation of the situation relating to a theoretical conception like that of 'objective' reality (which does not take into account limitations deriving from imperfections of ourselves and of the instruments with which we make our observations). It might even be said (from the opposite point of view) that it does not make sense, because things only acquire meaning through, and as a result of, observation and measurement. Either way, it seems that one is less open to criticism in starting from such an 'overdone' schematization of some 'objective reality' (that one may or may not take seriously), since this is merely the starting point for the introduction, as and when it suits us, of the gradual qualifications that take place in the transition from the metaphysical notion of 'objective truth' to the effective notion of 'verifiability'.

As we have already mentioned in Sections 1 and 2, these modifications and qualifications will have to be examined from various points of view, in relation to various circumstances and factors. As a first step, let us consider the *time* factor. In order to simplify the question, and to separate it off from others that are almost always connected with it, we restrict ourselves to the case in which, from a given instant t onwards, the result is known to everyone (or, at any rate, accessible, there being no need to do anything in order to ensure its occurrence or to learn about it). In order to fix ideas, think in terms of the entire output of news from the press, radio and television (political news, day to day items, weather, sport, the economy, science, the arts and so on).

Several cases may arise for each event E, according to the instant at which the result is known. An event may be *dated* more or less precisely, in the sense that the instant at which one knows whether it is true or false is known *a priori* (or, if we are dealing with a random quantity, the instant at which its value is known). For instance, the maximum temperature in Rome in August of next year (together with the fact of whether or not it is greater than that of the preceding August) will be known immediately after the end of the month in question. The population of Italy at the next census (the year of this being fixed by law, the actual day not yet decided) will be known shortly after the time mentioned (i.e. the year already fixed by law and the day chosen); and the same is true for the question of whether the population of the North or of the Central-South has had the greatest increase. It is easy to think of other examples.

In other cases, there exists a maximum time limit (possibly rigid, possibly not) before one knows the answer; and this advance knowledge could be relevant to only one form of answer (either affirmative or negative), or to both. As examples, consider the fact of an individual remaining alive, or continuing to hold his present post without interruption, or never being sick, or never having a car accident, and so on, up to some preassigned time or age. These are all statements that can only be ascertained as true at the last-mentioned time (if at all; for it could turn out that they were known to be false at some earlier instant). As for examples the other way around, it suffices to consider the negations of the above statements. If the preassigned time limit is dropped, and one stipulates 'until death', the conclusion[16] remains the same, except that a maximum time limit can no longer be given with certainty (although in practice the time at which a person would become 100 years old might be considered appropriate for the purpose in hand). The asymmetry between affirmative and negative answers vanishes if, always with respect to the life of a given individual, we consider examples like his dying because of an accident, or some other cause, or before (or after) some other individual, or in Italy, or abroad. (If in these examples one wishes to include a time limit, it is necessary to decide in advance which result is considered as valid if the limit is reached; for example, if the individual were insured against one of the two eventualities up to the time limit, to reach this limit would be equivalent to the occurrence of the other.)

There are cases, however, in which a statement can never be either verified or disproved until the end of time (or can only be verified or only disproved, or the one and the other, but it is not known if and when). Obvious examples of statements that can never be settled one way or the other within a finite time period are easily found; one only has to consider sequences of events unbounded in time. For example: tosses of a coin; spins of a roulette wheel; rainy and dry days (in a given locality); normal days and those days on which men turn into rhinoceroses;[17] male and female births (in chronological order, in some given town); passages of people in one direction or another through a gateway; and so on. Taking the coin tossing example, for convenience, it is sufficient to say, for example, 'neither Heads nor Tails will always occur', 'the frequency of Heads will tend to $\frac{1}{2}$', or 'will tend to some limit', or 'will exceed both the bounds 0·01 and 0·99 infinitely often', 'from some point on Heads and Tails will alternate in a regular fashion, HTHTHT…', 'the sequence HTTHH…, which represents the Divine Comedy in binary code, will be repeated an infinite number of times', and it would be easy to go on in this way. If instead we were to say that 'the frequency will be less than 0·01 at least once (after the first 1000 trials)', or 'we will have (after the first 1000 trials) at least one run of Heads as long as the preceding segment' (i.e., starting at the $(n+1)$th toss it reaches at least to the $2n$th), and so on, we would have examples of statements which, if true, are certainly verifiable within a finite period of time (although we do not know how long). For the complementary statements the opposite holds. If we add to each of the above statements something like '… or the frequency tends to $\frac{1}{2}$', we have statements which, if true, can either turn out to be verifiable within a finite time period, or not. An example of a statement that may or may not be verified within a finite time both if true and if false is the following: 'if one has, at least once (after the first 1000 trials), a sequence of identical results (always H or always T) from toss $n+1$ to toss $2n$, this will occur for the first time for the outcome Heads; if this never occurs, the frequency of Heads will have lower limit $\geqslant \frac{1}{2}$'.

16 Except for the first example, which becomes meaningless.

17 Up until now, all days have been 'normal' (Ionesco notwithstanding), and I think that this will always be the case. An example of this kind was needed, however, in order to make clear that it in no way differs from the others insofar as verifiability is concerned.

All the events we have considered, expressed as statements about some given sequence of 'trials' of unlimited duration in time, or concerning phenomena (like the death of an individual) that can take place at various times, can be more completely described by means of the ordered pair (E, T) (event E, taking values 0 and 1, and random time T, with $0 \leqslant T < \infty$, or $T = \infty$). This enables one to identify not only the truth value (True or False) but also the time (finite or infinite) after which the event turns out to be verified. A useful convention, enabling one to reduce the ordered pair to a single random quantity, is that of taking T with a + or − sign, according to whether E is true or false: that is putting $T^* = T.\,(E - \tilde{E}) = T.\,(2E - 1)$ (which, in fact, gives $+T$ if the time is T and $E = 1 =$ true, $2E - 1 = 1$, and gives $-T$ if the time is T and $E = 0 =$ false, $2E - 1 = -1$). We can summarize the various cases by expressing them in terms of this random quantity T^*. Both for positive values and for negative values (and independently of what happens in the other case) the possible values for T may either reduce to the unique value ∞, or to ∞ together with some finite values, or to finite values only (and in this case either unbounded or bounded; possibly 'bounded in practice', to put it in an unorthodox way).

What happens if we consider logical operations on events whose verification may be put off until different times, possibly 'never'? It is clear that under negation T^* changes sign; if $E_2 = \tilde{E}_1$, $T_2^* = -T_1^*$. If we consider the (logical) product, $E = E_1E_2$, E must turn out to be true if and when this has happened both for E_1 (at time T_1) and for E_2 (at time T_2), that is at time $T_1 \vee T_2$ (the larger of the two). In the opposite case, E will turn out to be false as soon as either E_1 or E_2 does; that is either at time T_1 or at time T_2, or at the smaller of the two if both events turn out to be false. All this can be condensed by means of the following convention: for $E = E_1E_2$ we have $T^* = T_1^* \vee T_2^*$ with the convention of modifying the general meaning of the sign as follows; 'take the larger of the negative values (the smaller in absolute value) and if they are both positive take the larger'. It is easy to see that this rule also holds if we include the values $+\infty$ and $-\infty$, and that it can be extended to deal with the product of an arbitrary number of events. By means of this rule, one can construct the set of possible values of T^* starting from those for T_1^* and T_2^* (and possibly others) *provided they are logically independent*.[18] Moreover, given that the logical sum is the negation of the product of negations, it is immediate, from what we have said above, that for $E = E_1 \vee E_2$ (and also for several events) we have $T^* = T_1^* \wedge T_2^*$, with a convention dual to the previous one: 'take the minimum of the positive values, and if they are all negative, again the minimum (i.e. the maximum in absolute value)'. On the other hand, these rules are obvious if we think of the meaning of the actual problems themselves.[19]

Actually, it is clear (and it also follows formally from what we have said) that logical combinations of events which are certainly verifiable within finite time periods yield events that are (in general) certainly verifiable within the most extensive of these time periods.

18 The meaning is the usual one: however, it might be useful to discuss it in our present context. T_1 and T_2 are logically independent if, when t_1' is a possible instant for E_1 turning out to be true, and t_1'' is a possible instant for it turning out to be false, and the same is true for t_2', t_2'' with respect to E_2, it is also possible that E_1 turns out to be true at t_1', and E_2 false at t_2'' (and similarly, interchanging true and false, for t_1'' and t_2', t_1' and t_2', t_1' and t_2').Of course, if logical independence did not hold, the set of possible values for T^* would either still be the one so determined, or a subset of it.

19 In order to avoid confusion, it would certainly be useful to introduce modified notation to replace $\vee$ and $\wedge$ if we planned on making actual use of it. In fact, we are only going to use the idea here, temporarily, for the purpose of this explanation: it is, therefore, not worth complicating things.

If, however, we consider an infinite number of events, which are each certainly verifiable within finite, although unbounded, time periods, we no longer have certain verifiability.

Let us now turn to the examination of the points raised by our analysis of the circumstances surrounding the notion of an event (with special emphasis on the realizability of a bet, which serves for us as something of a 'touchstone'). It seems natural to conclude that every postponement, and every asymmetric feature (between the ascertainment of the truth of either result) which might be caused by it, has an adverse effect on those characteristics required for an event until, if the postponement is too great, or even forever, these characteristics completely disappear.

Although we have to postpone the most relevant comments until after the introduction of probability, it is certainly clear, even at this point, that it would be very strange to discuss the truth of a statement, or bet on it (or even to maintain that it makes sense), when it does not assert anything that would enable one to discriminate in any way between possible future observations, even were one to think in terms of living for ever, or of passing on the task to future generations (assuming they never become extinct). And how silly it would be, besides being strange, to bet on a statement constructed in such a way that it is possible to lose right now, but to win only after the end of time-without-end (and this is so, even if it is a statement involving only a very small risk – like the assertion that 'the day on which men will turn into rhinoceroses will never come').

6 Verifiability and the Operational Factor

In Section 5, for the purpose of isolating the *time* factor, we restricted ourselves to considering the result as 'known to everyone (or, at any rate, accessible, there being no need to do anything in order to ensure its occurrence, or to learn about it)'. This assumption is not tenable as anything other than an idealized limit case, because even to listen to the radio, or to read a newspaper, requires some time, effort and cost, even if only a small amount. In general, however, it is necessary to do a great deal more in order to check the truth or falsity of an event or statement. It is often necessary to actually experiment in order to observe, or measure, or even produce, the phenomenon under consideration. In any case, even the mere recollection of existing data, or the task of researching for information concerning data already collected, can be involved, non-trivial operations.

We use the term *operational* factor to describe anything which, by virtue of the nature of such operations, imposes constraints on the verifiability of events. One aspect of this is the *cost* factor (intended in a broad sense), which we shall mention in conjunction with it. 'Precision' and 'indeterminacy' are other factors closely connected with the operational factor, but, because of their importance, we shall treat them separately in Sections 7 and 8.

In the first place, we come across a feature similar to that encountered when we dealt with the time factor: this is the difficulty, or even impossibility, of performing an excessive (or infinite) 'number' of operations (we say 'number' even though the terminology only makes sense in certain cases). For example, if we want to ensure that in a given time interval, $t_1 \leqslant t \leqslant t_2$, a certain quantity $y = f(t)$ has not exceeded a given level $y = y_0$, and for this purpose we wish to measure y (or just to check that $y \leqslant y_0$) at *each* of the infinitely many instants of the interval, the task would appear impossible to carry out. (And the same would be true even if measurements were only made at a dense subset of time points – also infinite, even if denumerable.) This statement is not, of course, to be taken as indisputable, deriving from some metaphysical prejudice, but simply as an empirical observation that seems undeniable in many practical instances (and in cases where it could be challenged we will acknowledge the fact). One cannot, however, conclude from this that there is no way of verifying the event under consideration. For the example in question, it suffices to invent and install some device like a 'max–min' thermometer, as used for temperatures, or like a fuse, calibrated to take an upper limit of electric current.

In the example we have considered, note the following two circumstances. Firstly, the constraints thaat derive from the impracticability of simultaneously considering an infinite number of trials might, formally, be the same as those of Section 5, but the significance is entirely different. In this context, it is not the *postponement* of the verification, *sine die*, which we are concerned with, but rather its *unrealizability* (were one able to do it, it could be done in a finite time). Secondly, the fact that a statement is not verifiable by means of some given procedure (here, measurements at an infinite number of instants) does not preclude its being verifiable in some other way. Unverifiability in some absolute sense cannot be asserted on the basis of the unrealizability of some or all of the operational schemes put forward so far: it can only be asserted on the basis of some rather general assumption that excludes realizability under *any* scheme (and the basis for such an assumption may or may not be very secure ...).

A formulation that might in certain cases adequately express the meaning of such constraints (albeit in a very schematized and idealized manner) could take the form of considering an event – or, better, a partition into events – as verifiable if one could reach it by means of a finite number of elementary, realizable operations. Note that if one regards the results of these operations as basic events this corresponds to assigning some special status to them – the very thing we strived so hard to eliminate! There is, in fact, no contradiction. In the first place, we have here made precise which criterion, if any, is to determine the basic subdivisions, and, hence, to assign them special status. Secondly, this would be justified only if in certain cases a formulation like the one we have just put forward as a hypothetical example appears to be actually valid (or, at least, almost so).

A similar, but more realistic, limitation (and not only from a practical point of view) would consist in restricting ourselves not simply to a finite, though arbitrarily large, number of operations but to a number not exceeding some given finite upper bound. In the case of the measurement of $y = f(t)$ in (t_1, t_2), for example, we must not merely take it to be impossible to make an infinite number of observations within the time interval, but also impossible to make more than some given finite number, which can be specified more or less precisely. In actual fact, this upper bound will, in general, be anything but precise unless we introduce the *cost* factor. Usually, one does not find a clear-cut point of separation up to which we can proceed without any difficulty but beyond which it is impossible to proceed. On the contrary, the fact of the matter is that one encounters ever increasing difficulty as one proceeds further and further. We use the term *cost* to denote the measure of these difficulties. As we have already remarked, it is a question of 'cost in a broad sense'; not simply the money spent but also the efforts made and the time required, taking into account the other alternative uses to which the time and effort might have been devoted.

We shall, however, always express this *cost* in monetary terms. This is done not so that we can adopt the economist's approach to the problem but simply to enable us to note that the problem of limitation can then be put in the following terms:

- First version: given the total budget available (over a certain time period), one can work out whether a certain sequence of operations is realizable or not. One must then select the most efficient from among those that are realizable (i.e. the one which gives the best overall result).
- Second version: this is a refinement of the first one and, following along the same lines, assumes that the best realizable sequence of operations is determined for a given total budget. The latter is no longer regarded as given and fixed, however, but as variable within some given range of values. In this version, the cost is also a matter of choice and will be chosen in such a way that one arrives at the equilibrium point in the neighbourhood of which an increase in cost produces an equivalent increase in efficiency (the principle of marginal returns).

The simplest assumption, so far as costs are concerned, is that of additivity (a given cost for each separate operation); in general, however, it is not necessary to limit oneself to this special case.

A further step towards realism consists in taking into account at least one particular kind of uncertainty which, in general, affects the 'operations' we employ in order to verify events (this is in fact the final such step we shall take: it seems to be adequate for the purpose of an idealized schematization). So far, we have assumed that such operations should always give us a precise answer, either YES or NO, to the question posed. We now make the assumption that the answer could also be MAYBE; in other words, either the experiment does not succeed, or it gives a result which is not sufficiently clear-cut to enable us to consider either YES or NO as established beyond doubt.

It is commonplace to remark that this can happen in any kind of experiment or procedure (such as the examination of a witness's testimony in court). If there is any well-founded objection to this statement, it would be that admitting MAYBE does not go far enough and that a clear-cut and definitive YES or NO cannot be obtained from any experiment … and that it is always a question of greater or lesser probabilities. It goes without saying that I very much agree with this but, in order to avoid getting into a vicious circle, it seems to me that the best one can do, or, at any rate, the least objectionable alternative, is to go along with this approach. It introduces uncertainty in a meaningful way, by including the possibility of the answer MAYBE, but does not preclude the construction of a scheme coming before the introduction of the notion of probability. Such a preclusion would arise if the woodworm of uncertainty were to find its way into the answers YES and NO. On the other hand, a study based on a formulation that was completely rooted in uncertainty (that is to say probability) could be carried out once probability theory was constructed. This is done, for example, in the classical theory of errors, where an error can be arbitrarily large, although with an extremely small probability (given by the normal, or Gaussian, distribution).

With respect to an event E, an operation may, therefore, yield either the answer YES (and hence NO for $\tilde{E}$), or the answer NO (and hence YES for $\tilde{E}$), or the answer MAYBE (the same for both E and $\tilde{E}$). But it may very well happen – and this is the case we shall consider – that the operation yields similar answers for other events, and ultimately for a partition, $C_1, C_2, \ldots, C_s$. In this case, the following possibilities arise: either the answer

YES is given for one of the constituents C_k (and hence NO for all the others – a complete answer); or there are no YES answers and more than one MAYBE, with or without NO answers (all MAYBE answers give an absolute void, leaving one in the same state of ignorance as before; if there is at least one NO we have a partial answer). With respect to the event E under consideration, the answer will be YES if E contains the constituent with the YES answer, or (if YES was missing) if it contains all the constituents with the answer MAYBE; the answer will be NO if, symmetrically, E is contained in the union of the constituents with NO answers; it will be MAYBE otherwise (i.e. if both E and $\tilde{E}$ are compatible with the union of the NO constituents and with that of the MAYBE constituents, replaced by the YES constituent if it exists).

Taking account of the partial answers in this more detailed manner – that is referring to the partition into constituents – greatly increases the efficiency of the procedure, because it permits us to draw the maximum possible information from the result of each operation. As an intuitive illustration of this obvious mechanism, suppose that in picking out the guilty party from among n individuals (who constitute the entire collection of possible suspects) we obtain $n - 1$ observations having the very limited effect of excluding one with the answer NO, and attributing MAYBE to all the others. If each single observation excludes a different individual, the unique remaining person must be the guilty party; if instead one had been concentrating on the guilty party (perhaps because he was the main suspect) and had only recorded whether or not each trial was sufficient, one would have had to conclude MAYBE, having always obtained the answer MAYBE.

So far as the applications to verifiability are concerned, this enrichment of the possibilities makes life somewhat more complicated, but compensates for this by removing some of the other complications deriving from the rigidity of the previous scheme. We shall see this particularly when we come to deal with measurement procedures (in Section 7).

On the other hand, everything could be expressed in a more direct and straightforward way by simply referring to the field Q of elementary cases Q (rather than fixing one's ideas – as, in a certain sense, is more instructive – on that subject which the operation has brought into play). Thinking in terms of Q, an 'operation' has the intended purpose of obtaining information, which means, in fact, narrowing the field Q by eliminating the Q that have turned out to be impossible (NO), and retaining the others (MAYBE). The YES would only be of use in the case of complete information, in order to pin-point the unique nonexcluded 'point' (or never, if one takes strict account of the observation that no subdivision – even if words like point or atom are used – can ever be considered as definitive, or as the ultimate one).

So far as the previous considerations about the advantages deriving from efficiency and cost are concerned, the only difference lies in the fact that the factor of uncertainty is introduced. If we decide to proceed in some given manner, it is no longer known *a priori* whether, and after how many operations, we will reach the desired conclusion. To judge a procedure as appropriate therefore necessitates the application of the theory of decision making under conditions of uncertainty – that is the maximization of expected utility corresponding to an uncertain cost and uncertain efficiency.

The experimenter could, taking everything into account, estimate, on a rough and ready basis, the procedure that he thinks to be the best. If he decides to apply the calculus of probability, so much the better (provided it is worthwhile to do so; i.e. that the additional cost of performing the calculations does not exceed the expected increment in utility for choosing the optimal procedure). In any case, in this context we are not drawn into the vicious circle we previously indicated that we wished to avoid. The possible application of the theory of probability to this aspect of the problem is something which concerns the experimenter; he, independently of the stage in the present treatment at which we have occasion to speak of him, might equally well know or not know the calculus of probability. On the other hand, the fact of whether he has worked things out for himself well or badly (i.e. of whether he chooses the procedures in a more or less advantageous manner) is something of no concern or interest to us. Here, we are only interested in procedures in the abstract, in principle, as instruments we can make use of, and which give certain types of answers. Whether they are used well or badly is a separate question.

7 Verifiability and the Precision Factor

We have already (in Section 6) made some mention of measurements but, in order to divert attention from the topic being considered, we tacitly assumed that we were dealing with exact measurements. It is well known, however, that exactness is unattainable (except in counting procedures – provided one makes no mistakes). When dealing with measurements, one can only proceed by fixing in advance some higher or lower level of accuracy or *precision*.[20] Here, too (as in the case of similar questions considered previously), an improvement in precision generally implies an increase in cost. Looked at from this point of view, there is nothing new, and nothing to add to the above.

The most important issue, which we must examine, concerns the implications of imperfect precision in the measurements for the identification of the individual points Q of $\mathcal{Q}$ (or of $\mathcal{S}$). This leads into a discussion of whether, and in what sense, it is appropriate to introduce a topological structure into the field $\mathcal{Q}$ (or $\mathcal{S}$). We have, of course, struggled hard to eliminate any trace of such structure – as in the case of the special status of basis events – but here, as we found in that case, too, there is no contradiction. It will simply be a question of rejecting structures that are either unnecessary or suggested by so-called motives of analytic convenience and possibly accepting, after careful examination, those structures that correspond to essential and meaningful requirements.

Let us begin by considering the case of a single random quantity X, having the real line as its set of possible values. Thus far, in beginning our discussion of 'verifiability', we have considered events rather than random quantities. It is clear, however, that 'to verify' the value x taken by X is (precisely) to verify which of the events $E_x = (X = x)$ is the *true* one: in a weaker sense, it would be a question of verifying events of the form $E_I = (X \in I)$, with I, *a priori*, arbitrary.

There is nothing to be said on a general level about the possible ways of proceeding. This is not a mathematical problem, but rather one of the peculiarities inherent in each individual case. There is no problem when X appears already in a simple form, written in Arabic numerals (like the number on the ball at bingo, or on the sectors of the roulette wheel), or spelt out in the form of dots (as on a die), or is provided and vouched for by

20 In probability theory (and especially in relation to 'error' theory) the term 'precision' is often used to denote the reciprocal of the standard deviation, $1/\sigma$. Here, however, we use the word informally, without reference to any 'technical' meaning (as in Chapter 12, 12.3.1).

others (like the data from a census, or statistical data in general). In these cases we are dealing with integer random quantities and this, of course, is a simpler situation. It is possible, however, to find similar, immediate and rather precise representations when dealing with continuous quantities (physical constants, geographical coordinates, geodetic points, heights of buildings, weights of objects and individuals etc.). If, on the other hand, it is our responsibility to pin-point (exactly, or in part, or approximately) the value x of X, we have to examine, case by case, which events E_i are more or less accessible to us, and which of them yield the more interesting or useful pieces of information about X. Having done this, we then choose (as in the case already noted in Section 6) that combination of operations which is most advantageous, taking into account considerations of both efficiency and cost.

In what way is it possible to learn something about a random quantity X? And what can we learn? Posed in these terms, the question does not make sense because the answer does not depend on the fact of X being a random quantity, but rather, case by case, on the actual and particular meaning attached to each given X by virtue of its definition. The case usually dealt with, to the virtual exclusion of all others, is that of a physical quantity for which one may obtain more or less precise measurements; this will also be our principal concern here. It should be noted, however, that, apart from this rather special case, there is no reason to think that, in general, the problem can be posed in terms of the same concepts. An X defined as a (practically speaking) insufficiently continuous[21] function of another random quantity (having physical meaning) clearly does not lend itself to measurement through such a definition; but it may be measurable by virtue of the fact that it, or some suitable function of it, has a physical meaning of its own, which renders it capable of direct measurement. On the other hand, it may happen, especially if the definition of X is bound up with mathematical concepts, that the question of whether X belongs to sets I which are less straightforward (and in general less resolvable) than the intervals turns out to be easily answerable, or, at any rate, feasible (e.g. one might ask whether or not X is rational, or algebraic etc.).

The random quantity $X = \pi$ provides a suitable example. This might seem rather strange because – it might be argued – π is *not* a random quantity; it is a well-determined number, already determined with remarkable accuracy (given his time) by Archimedes, and for which explicit expressions in the form of series have been discovered, so that it is now known 'more precisely than any other number'.[22] Agreed: but π is not known (and is thus random, since we do not accept a more restrictive use of the term) insofar as its remaining decimal places are concerned; just as, for a long time, it was not known whether it was rational or algebraic. These questions, as everyone knows, have now been resolved (negatively): it is clear that this would not have been achieved merely by the more and more precise determination of the numerical value of π – even by going on to infinitely many decimal places – in order to look for possible periodicities (supposing it were rational) and so on. Instead, there was another approach (just as in the previous example concerning the maximum of a function in an interval).

Another example (of a simpler but more artificial kind) is provided by the following. Suppose we define $X = (N + 1)(N < \infty) - 1 + E'/4 + E''/\pi$, where π is no longer considered as a random quantity, E' and E'' are arbitrary events (e.g. a sports result and some feature

21 For example, in the sense of there being a Lipschitz condition, $|f(\xi_2) - f(\xi_1)| < K|\xi_2 - \xi_1|$, with K not too large (at least in the range of practical interest).

22 To 100 000 decimal places; given by D. Shanks and J. W. Wrench in *Math. of Computation* (1962).

of the weather), and N is the number of (possible) exceptions to Goldbach's conjecture.[23] The first two terms in the expression for X simply denote N, to be replaced by -1 if it is infinite; it is, therefore, an integer if the last two terms are missing (E' and E'' are false: $E' = E'' = 0$), rational (noninteger) if E' is true and E'' is false ($E' = 1$, $E'' = 0$; $\frac{1}{4}$ is added), and irrational (in fact, transcendental) if E'' is true ($E'' = 1$, it does not matter whether $E' = 0$ or $E' = 1$; either one adds $1/\pi$ or one adds $1/\pi + \frac{1}{4}$). In this example, it is therefore sufficient to know the outcomes of E' and E'' in order to establish whether X is integer, rational and noninteger, or transcendental; on the other hand, one would need to know whether Goldbach's conjecture was true in order to know whether $0 \leqslant X \leqslant 1$; or whether there were an infinite number of exceptions in order to know whether X were negative ($-1 \leqslant X < 0$); or a finite number (N) in order to know whether $X \geqslant 1$ (more precisely, $N \leqslant X < N + 1$).

Having put forward these examples and discussed them, we can now confine ourselves to cases where X has a physical meaning (or something akin to a physical meaning). To be more specific, we shall deal with those X for which knowledge of their values can only be attained or approached through operations of measurement (which may or may not be precise). A few illustrative examples will make this intuitive explanation clearer; it, in turn, will prove useful for a careful scrutiny of the same ideas insofar as they constitute the standard formulation. We are interested in delving deeply into certain aspects of the latter, but we are even more concerned to warn against the customary readiness to accept that these assumptions can, or must, be assumed valid for 'all' random quantities, with no discrimination. The examples and discussion given above were intended precisely for this purpose.

In order to provide a technical discussion of this topic, we must first of all say what we mean by an 'operation of measurement'. By this we mean an operation which, if applied to X, can only have the effect of restricting the field Q of possible values to an interval[24] (and we repeat that this is an 'experimental' question, not a mathematical one).

These general characterizations only serve as a starting point. It will prove much more important for our analysis to examine closely a few of the main variants from among the many possible. As far as *precision* is concerned, it is of interest to single out the three cases of *bounded, unbounded* and *perfectible* precision; and, moreover, cases of precision which are *possibly perfect* and *certainly perfect.* So far as *partitions* are concerned, we shall distinguish the three cases of *fixed*, *free* and *variable* partitions. It would be pointless and rather tiresome to examine all these variants and their innumerable subcases in detail, not to mention all the other possibilities. What is worthwhile is to exemplify some of them, in order to single out the points which really matter (noting in passing that in this way we obtain an adequate view, even if not complete).

We shall begin with a typical example of a measurement operation with bounded precision and a variable partition. We obtain a value which we know to differ from X by not

23 Which asserts that every even integer can be written as the sum of two primes. Investigation has (so far) revealed no exceptions, but no proof has yet been found, so that there could well be exceptions (even an infinite number of them).

24 It is convenient to fix one's ideas on this case (that of a *two-sided* restriction); it is better, however, not to exclude the limit cases in which the interval becomes a half-line (*one-sided* restriction), or the whole line (*no* restriction; operation failed); or, instead, going to the opposite extreme, into a point (exact measurement). It goes without saying that if $\mathcal{Q}$ is not initially the whole line then the above reduce to the intersection of $\mathcal{Q}$ with the interval in question (and the operation may have no effect if the actual restrictions obtained are weaker than those already holding for $\mathcal{Q}$).

more than a given maximum error, δ. (One could consider, asymmetrically, different limits, $-\delta'$ and $+\delta''$, for negative and positive errors, or one could think of δ – or even of δ' and δ'' – as functions of x, where $X = x$; one could complicate matters even further, by thinking of other random quantities, and so on; however, nothing of conceptual importance is lost, and much is gained in clarity, if we limit ourselves to the simplest case.) Having obtained, by measurement, a certain value x, it tells us that X certainly belongs to the interval $(x - \delta, x + \delta)$. The precision is *bounded* in the sense that we know X up to this δ, whatever it may be. If, however, we have at our disposal operations of this type with δs arbitrarily small, then the operation which consists in first selecting one of them, with a δ corresponding to the degree of precision required, and then in performing a measurement on the basis of it, is an operation with *unbounded* precision (there is always some preassigned margin of error, δ, but this can be chosen arbitrarily small before the measurement is performed).

Even with a fixed δ, we could perform measurements of *possibly unbounded* precision (in the sense that the margin of error might remain equal to δ, but, by chance, could also become arbitrarily smaller). Bearing in mind that all 'assumptions' (like the ones we are about to make) always have an experimental interpretation – even if this is not explicitly mentioned – it is sufficient to assume that it is possible to 'repeat' the previous operation; or, to put it in a better way, to perform another operation with identical characteristics (and we are assuming that there are no constraints of a physical nature, or of cost, administrative veto etc.). If this is done, and two distinct values $x_1 \neq x_2$ are obtained (to fix ideas, suppose $x_1 < x_2$), the two bounds reduce to the single interval

$$x_2 - \delta < X < x_1 + \delta,$$

so that X belongs to the interval

$$x \pm \delta^*, \quad \text{where } x = \tfrac{1}{2}(x_1 + x_2) \text{ and } \delta^* = \delta - \tfrac{1}{2}(x_2 - x_1).$$

Observe that the same conclusion (with more and more possibility of obtaining a small value of δ^*) holds in the case in which one can repeat the same operation over and over again, as many times as one wishes. It is enough to take x_1 to be the minimum and x_2 the maximum of the measurements obtained. Put into words: if one takes the mean of the resulting maximum and minimum, one has a margin of error equal to that of a single observation (δ) less half the difference between the maximum and minimum. Assuming that one could go on to make an *infinite* number of repetitions, the precision is even *possibly perfect* (it suffices to take x_1 and x_2 as the infimum and supremum rather than the minimum and maximum) because it is possible that the difference $x_2 - x_1$ tends to 2δ and, hence, δ^* tends to zero.[25]

25 This could also happen with only a finite number of repetitions, even with just two, if one assumes that the intervals $x \pm \delta$ are to be taken as *closed* (i.e. that the error can reach its upper bound δ). The question is, in itself, rather hair-splitting (given that to have fixed a *precise* δ is already an arbitrary schematization), but it is precisely for this reason that it seems innocuous to follow here the criterion of 'mathematical convenience' (which, in general, we do not approve of). From this viewpoint, it appears preferable to take the intervals to be *open* (and that is why we chose to write them, without comment, in the form $x - \delta < X < x + \delta$, with '$<$' and not with '$\leqslant$'), thus avoiding the fact that the conventional assumption of a precise separation leads (even though exceptionally) to the possibility of an exact measurement. If the intervals are open, every intersection of a finite number gives an open interval, and hence not a point (neither can it be empty, since we cannot have $x_2 - x_1 > 2\delta$).

If, finally, one proceeds to 'repetitions' of the operation with different precisions, the conclusion does not change in the finite case, nor does it (formally) in the infinite case. The bound can always be written

$$x' - \delta' < X < x'' + \delta'',$$

with $x'' + \delta''$ equal to the minimum (or infimum) of the $x_i + \delta_i$ (and vice versa for $x' - \delta'$). We note, however, that, because the δ_i are also variable, the minimum (and maximum) will no longer be attained, in general, for the i which give, respectively, the minimum and maximum x_i. The interval $x \pm \delta^*$ will be centred at

$$x = \frac{1}{2}(x' + x'') + \tfrac{1}{2}(\delta'' - \delta') \quad \text{with} \qquad \delta^* = \tfrac{1}{2}(\delta' + \delta'') - \tfrac{1}{2}(x'' - x').$$

It is instructive to note that, in any case, δ^* is at most equal to the minimum (or infimum) of the δ_i, given that the interval $x \pm \delta^*$ is contained (as their intersection) in all the intervals $x_i \pm \delta_i$. If, therefore, we have at our disposal operations with δ arbitrarily small, and we can not only choose one (as in one of the cases mentioned above) but can also perform a succession of them with $\delta \to 0$ (or in some way with infimum zero), we have a case of *perfectible* precision if one thinks in terms of performing a finite number of operations, however large (or if one takes into account that at any time point not infinitely far away the number will be finite anyway). We have a case of *certainly perfect* precision if we place ourselves at the end of time, or if we imagine that an infinite number of operations can be performed in a finite time.

We have mentioned precision in its various forms but as yet we have said nothing about the partition. We have merely noted that, in the example considered (and extended in various ways), we had a *variable* partition, since the intervals which could be obtained as bounds could have any end-points whatsoever (with no imposed constraints inherent in the nature of the operations considered). The opposite case – that of a *fixed* partition – is perhaps best illustrated in the case of a number X for which (as in the example of $X = \pi$) one can determine the decimal places one at a time (or take readings from a series of increasingly finely graduated scales). A *free* partition is a partition that one is free to choose from among several others (or, in general, from among an infinite number), but which, after the choice is made, becomes fixed. Think of a measuring scale whose origin and measurement unit can be chosen arbitrarily (for example, instead of X, consider measuring or calculating the successive decimal places of $\sqrt{2} + \pi X$; or measuring angles having rotated the dial of the instrument; or, more generally, to deal in this way with nonlinear functions – for instance, log X, or the tangent of the angle – and so on). A further example occurs with operations for which one can establish in advance the interval to which the question will refer (is it or is it not true that X is in the interval (x', x'')?). If one assumes (certainly) perfect precision, there is nothing else to be said. In general, however, this would not be valid (and, in any case, this is not the form of greatest interest). One can think, for example, that (because it is outside the scope of the apparatus) there may be some doubt as to whether X is near (inside or outside) the end-points, and, in this case, one can imagine that the answer is either YES or NO (by chance), or MAYBE. Supposing that δ is the margin of error (for different δs the same qualifications as were made previously continue to hold) we have the following: in the case of the two answers YES or NO only,

YES means that $x' - \delta < X < x'' + \delta$ and
NO means that $(X < x' + \delta) \vee (X > x'' - \delta)$ (in particular, it is *void* if $x'' - x' < 2\delta$).

In the case of three answers, YES, NO, MAYBE, we have

YES: $x' + \delta < X < x'' - \delta$,
NO: $(X < x' - \delta) \vee (X > x'' + \delta)$,
MAYBE: $(x' - \delta \leqslant X \leqslant x' + \delta) \vee (x'' - \delta \leqslant X \leqslant x'' + \delta)$.

It is perhaps more interesting to consider the simplest case, that of comparison with a single value x, asking whether $X < x$ (the usefulness of this particular case is obvious if one considers that an interval corresponds to two questions of this kind with $x = x'$ and $x = x''$). In this case, the answers have the following interpretations: with the two answers YES or NO,

YES: $(X < x + \delta)$
NO: $(X > x - \delta)$.

(In other words, whatever the answer is, there is no way of excluding the possibility that X can assume any of the values in $x \pm \delta$). On the other hand, with three answers, the MAYBE clearly picks out this possibility;

YES: $X < x - \delta$ (and, *a fortiori*, we certainly have $X < x$),
NO: $X > x + \delta$ (and, *a fortiori*, we certainly have $X > x$),
MAYBE: $x - \delta \leqslant X \leqslant x + \delta$.

In both cases, however, it could happen that the answer MAYBE does not give any information. This would happen, for example, if it could arise both as a result of X being near the end-points of the interval under consideration, and as a result of a chance malfunction of the apparatus. We make this comment not for the pleasure of adding yet another variant, but in order to stress that every question relates to factual circumstances. The logical structures can only be specified with respect to the given circumstances, and not on the basis of some convention fixed beforehand in the belief that one can do without taking these circumstances into account.[26]

Let us now consider the conclusions we can draw as a result of the above discussion. If X is a random quantity, knowledge of which can be attained by means of procedures of the kind considered above, for which sets I can one say that the event $X \in I$ is verifiable? In other words, when is it meaningful, practically speaking, to ask whether X belongs to I?

We have throughout, when introducing a notion, tried to give a realistic analysis of the underlying assumptions (bearing in mind, however, that it would be unrealistic to take these realistic considerations too seriously, thus turning them into metaphysical obsessions). Following this procedure, it would appear in this case that we are justified

---

26 My insistence on these points (which I hope the reader will excuse) is a consequence of my impression that in general they are not taken sufficiently into account (and sometimes not at all).

in confining ourselves to consideration of just three types of answer. We can label them by again using the terms *bounded*, *unbounded* or *perfect precision* (avoiding the introduction of other terminology, although, for the reasons already given, the trichotomy ignores many further subtle subdivisions).

If we confine attention, for the time being, to the case of a random quantity (geometrically, a straight line – the real line), the three cases can be described in the following way, beginning with the last one.

Perfect precision. This is the case in which one leaves out of consideration the practical difficulties examined in Section 7 and considers the value of X, and hence the question of its belonging to some set I, as perfectly determinable. This is very much a theoretical view but there is no reason to rule it out on these grounds. Indeed, there would seem to be no justification for closing one's mind against any formulation (even those which themselves would lead one to close one's mind against others), unless, in this case, one were taking account of the unrealistic character of such perfect precision, and, for this reason, were to choose to substitute one of the following cases in its place.

Unbounded precision. This is the case in which, by translating into a necessarily idealized form the practical situation that can be realized in the most favourable circumstances, one imagines that, apart from the endpoints, an unambiguous answer can always be obtained to the question of whether or not X belongs to an arbitrary set I. In other words: the question 'is it true that X belongs to the set I?' is partially decidable, in the sense that one can obtain either the answer YES or the answer NO; certainly YES if X is in I, certainly NO if it is not, but it could be either answer, YES or NO, if X belongs to the boundary, $\mathscr{F}(I)$, of I (that is, if every neighbourhood, however small, having X as an interior point, contains both points of I and of its complement I). Put another way, having obtained a *measurement* $\hat{x}$ of X, it will be legitimate to conclude that $X \in I$ if $\hat{x}$ is an interior point of I (and so on).

Bounded precision. This is a weakened form of the previous case, obtained by substituting in place of the boundary, $\mathscr{F}(I)$ of I, 'a boundary strip of width δ', which we denote by $\mathscr{F}_\delta(I)$, consisting of the points x for which the neighbourhood $x \pm \delta$ contains both points of I and of its complement. If we call δ-internal (δ-external) those points which belong to I ($\tilde{I}$, respectively) but not to $\mathscr{F}_\delta(I)$, the difference between this and the previous case reduces to using the prefix 'δ-' (we could perhaps even use the term 'δ-boundary'). One could even include the previous case in the present one by introducing a small modification and allowing $\delta = 0$: it would suffice to redefine $\mathscr{F}_\delta(I)$ by replacing 'the neighbourhood $x \pm \delta$' by 'every neighbourhood $x + \delta'$, with $\delta' > \delta$'.

In all cases, this must be interpreted as a possible final result (and not as a possible result of 'an operation' of measurement, which could be improved on by combination with others).

Bounded precision: more general forms, fixed and optional. It is not difficult to weaken this scheme in two different senses, thus obtaining a much closer correspondence to realistic requirements (albeit idealized ones).

Instead of neighbourhoods $x \pm \delta$, with δ fixed, we can define directly $\delta(x)$ as a neighbourhood of x associated with x in an arbitrary way (in general asymmetric, and of variable length), except for the necessary restriction of reciprocity; $x' \in \delta(x'') \leftrightarrow x'' \in \delta(x')$. For example, we could have neighbourhoods defined by $f^{-1}(f(x) \pm \delta)$, with f increasing, δ constant. This frees us from any particular scale.

A more substantial and indispensable freedom, however, is that we can consider the possibility of obtaining a measurement with precision defined by different laws, $\delta(x)$, a free choice being allowed among them (although – but for the time being we shall not worry about this point – the cost factor may differ). It might be that we are faced with a choice between measuring, with an error $\leqslant\delta$, either X or e^X, or one of three quantities, ..., or an infinite number, or X with an error δ (small, but not necessarily arbitrarily small). In such cases, we shall speak of measurements with *optional bounded precision*.

For each $\delta(x)$, the definition of $\mathscr{F}_\delta(I)$ becomes 'the union of all the $\delta(x)$ with $x \in \mathscr{F}(I)$'.

Expressed in this way, it becomes clear how, formally, the notions of the preceding schemes could be directly transported from the one-dimensional case (only a single X under consideration) to the general case (provided that we have a topology). It is, however, necessary to carry out a critical examination – not in the abstract, but from the point of view of the basic issues involved – of whether, and how, essential considerations about observability, similar to those presented above for a single X, can justify the introduction of a topology in the general case. (We did, after all, strive hard to eliminate any trace of topologies which might have arisen naturally, but perhaps, we suspected, irrelevantly.)

8 Continuation: The Higher (or Infinite) Dimensional Case

In the finite-dimensional case the extension of the considerations previously encountered does not involve too many new features. But when do we have a finite-dimensional situation?

Having reshuffled all the points in order to eliminate the topology, the S_r formed by the r random quantities $X_1 \ldots X_r$ *is* merely an infinite collection of points with the cardinality of the continuum, which can be individuated by means of a single number X. (It is not even necessary to have recourse to the Peano curve in order to obtain continuity, since we leave this out of consideration.) We are certainly not saying anything new if we point out that in eliminating the topology we take away the meaning from dimensionality, too; this is a necessary observation, however, if we are to frame the problem as it presents itself in our case.

Let us now make a point in the opposite direction, in order to show that S_r (assuming that we have adopted it in order to have a continuous representation with respect to $X_1 \ldots X_r$) might not be sufficient. If we were also interested in considering the values of another random quantity X, a function of the others, $X = f(X_1, X_2, \ldots, X_r)$, we do not need – logically speaking – a new dimension in order to represent it. If, however, the function f is very irregular – for example, everywhere discontinuous – knowledge of the X_h in the sense of 'unbounded precision' is not sufficient to determine the value of X. In the case of bounded precision, the same difficulty would arise even if X were continuous but varying sufficiently rapidly (for example, $X = \sum_h \sin \lambda X_h$, with $1/\lambda$ small in relation to the imprecision in the measurement of the X_h[27]). In these cases, if one is also interested in X (and if there is some way of measuring it which is more direct and dependable than calculating it by means of the formula – if not, there is no real interest, only a vain desire), it will be necessary to introduce another dimension for X, in this way going from S_r to S_{r+1} (and so on, in the same way, should we wish to consider several such functions).

Here is another observation, again of a practical nature, which, under different assumptions, reduces the number of distinct dimensions to be considered. Suppose that in the initial formulation one considers a large number of quantities $X_1, X_2,\ldots, X_r$, like, for instance, the coordinates and velocity components of N molecules (in which case, $r = 6N$). Suppose, also, that, from a practical point of view, we are interested in, and can measure, not the individual X_h (although this would be needed for the theoretical part of the development and calculations) but only some of the macroscopic quantities which derive from them (or, at any rate, some number of quantities r', much smaller than r). In order to study this aspect of the problem, one must refer not to S_r (having the number of dimensions required by the theoretical formulation), but rather to $S_{r'}$ (the space of the r' quantities, X'_k, $k = 1, 2,\ldots, r'$, functions of the original X_h, $h = 1, 2,\ldots, r$, which in practice are either unobservable or irrelevant: $X'_k = f_k(X_1,,,X_2,,,\ldots,,,X_r))$.

It appears, then, that the number of dimensions is also a notion that is neither absolute on the one hand, nor arbitrary on the other. It boils down to being the number of quantities that one requires to measure independently in order that one may know all the quantities of interest to the degree of precision judged satisfactory. Naturally, 'satisfactory' cannot have the meaning of 'to the desired extent' if such a degree of precision is not attainable. The term is then to be understood in the sense of 'we can be content with this, given that other methods of measurement (for example, direct ones) would not increase the precision to the extent required for it to be worthwhile (in terms of efficiency and cost) applying them'. All this depends, of course, on the kind of precision that is attainable.

This 'definition' of the number of dimensions (or, to put it in a better way, this 'suggested way of choosing it') is close in spirit to the necessary extension to this case of topological and allied notions. If by 'a small neighbourhood of a point' (permit me to use the expression) one means, in practice, a 'set of points indistinguishable from it', 'sufficiently close to it compared with the degree of imprecision of the measurement', this indistinguishability must be attributed to the measurements to be made in practice according to the above-mentioned requirements (even if this is a bit extreme compared with the rigid theoretical scheme of some previous formalistic approach).

This having been said in general, there is nothing to add in the case of unbounded precision (any definition of the neighbourhood of a point will do: for example $|X_h - x_h| < \delta(h = 1, 2,\ldots, r)$, or $\sum(X_h - x_h)^2 < \delta^2$). In the case of bounded precision, however,

27 The following provides a suitable, simple example. Suppose we are dealing with a phenomenon whose behaviour as a function of time t is given by $f(t) = A(t) + B(t) \sin \lambda t$, where $\sin \lambda t$ represents a daily or annual variation (with period equal to one day or one year), whereas $A(t)$ and $B(t)$ have slower variations (only perceptible on a much longer timescale) because they represent the trend of the mean value of $f(t)$ and the amplitude of its oscillation. In order to investigate its behaviour, it is certainly more instructive to consider and to represent $f(t)$ as a function of two variables, t and $\tau = t - 2\pi k/\lambda$ (where k is the largest integer for which $2\pi k/\lambda \leqslant t$): we then have $f(t, \tau) = A(t) + B(t) \sin \lambda\tau$. (And one day of time τ on the ordinate might appear larger than a year or a century of time t on the abscissa.)

it is necessary to say something more; especially in the *optional* case, which can give rise to a greater variety of circumstances.

In the case of fixed, bounded precision we still have a law δ which to each point $Q = (x_1, x_2,\dots, x_r)$[28] associates a given neighbourhood $\delta(Q)$ (with the condition that $Q' \in \delta(Q'') \leftrightarrow Q'' \in \delta(Q')$). We shall say that the measurements on the different X_h are 'independent' if the neighbourhood $\delta(Q)$ is the Cartesian product of the neighbourhoods $\delta(x_h)$. If the measurements are not independent, or are performed on combinations of the X_h, or in some other way, the form of $\delta(Q)$ may be anything at all.

This fact, which if δ is fixed has no practical importance, becomes important in the optional case where one can choose, from among various laws, $\delta_1, \delta_2,\dots, \delta_s$, the δ one prefers (it may be a choice from an infinite number of laws, $\delta \in \Delta$). This possibility of choice opens up the way to much more varied and important features in the higher-dimensional case and it would be pointless to attempt to give a typical example. It will be enough to present a case that throws light on a situation of particular interest, one to which we shall return later from another viewpoint. If X_1 and X_2 are 'complementary' quantities (in a quantum theoretic sense) there exists a relation of a probabilistic nature between the precisions with which one can measure them simultaneously. Translating this (in order to include it within the framework of the present considerations) into a relation which gives bounds, we could say that the margins of error, δ_1 and δ_2, of the measurements of X_1 and X_2 can be chosen at will subject to the restriction $\delta_1\delta_2 >$ constant. As a neighbourhood $\delta(Q)$ (in the (x_1, x_2)-plane) we have the option of any rectangle of constant area (i.e. with vertices lying on a rectangular hyperbola). So far as the question of the number of dimensions is concerned, this circumstance reveals possible causes of uncertainty and complications over and above those already mentioned. In fact, if we consider as adequate the knowledge of X_1 and X_2 with the precision attained by measuring them jointly, we are assuming that we can measure two quantities. If, on the other hand, we are not content with this precision, we can obtain an acceptable measurement for only one of the two quantities. We could choose which one, but it would be only one. Should we, therefore, eliminate the other from the set of quantities to which there corresponds a dimension? The question is (in all probability) just a rhetorical one, because if we do not eliminate the dimension in question it should not cause any extra trouble. Moreover, it is capable of either being excluded or being chosen, and hence it might be called 'potentially observable'. In any case, even these points which we leave open show (to paraphrase Shakespeare) that many more problems arise when we worry about what is 'realistic' than arise if we accept some pre-packaged scheme.

If we wish to go on to the *infinite-dimensional* case, we must again ask ourselves what it means. Again, there does not appear to be a unique answer. We might interpret it in a *weak* sense (thinking of being interested in representing an infinite number of quantities, but only being able to measure an arbitrarily large finite number of them, or combinations of them); or in a *strong* sense (thinking in terms of being in a position to say something which depends on the infinite number of quantities taken as a whole: e.g. to refer to the lim sup etc.).

28 We are using r here to denote the 'chosen' number of dimensions; it may be the theoretical r, or any other.

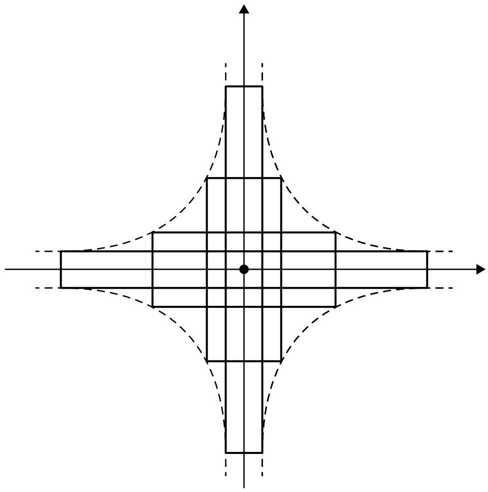

Figure A.1 Rectangles of equal area (concentric, and similarly placed); equilateral hyperbolae, the locus of the vertices.

This latter possibility seems rather theoretical, especially given the discussion of Sections 5 and 6: recall, for example, the discussion concerning the maximum of a function (this was judged to be indeterminable if an infinite number of measurements of all the values assumed were required and determinable if a direct method existed).

The weak interpretation seems to correspond better to the requirements of a theoretical scheme in line with the interpretations conceivable in real terms in practical applications. We shall, as a general rule, stick to this interpretation. For questions where the links with empirical considerations are so tenuous, it is difficult to base one's judgements and decisions on precise reasoning, rather than on impressions. Perhaps the choice here has been dictated by the idea that the weakest formulation is always the most valid one … until such time as it is shown not to be (and, in this case, one will always be able to patch it up after due consideration). The remedy would be more difficult in the opposite case.

Let us recall what is meant by a weak topology in an infinite-dimensional linear space. It is sufficient to require that among the neighbourhoods of a point are the half-spaces containing it. It follows, in fact, that the intersections of a finite number of them are also neighbourhoods (and this applies, *a fortiori*, to sets containing these intersections), and this completes the enumeration of the neighbourhoods. In terms of convergence, this is equivalent to defining the (weak) convergence of a sequence of points Q_n to a point Q by the condition that every linear coordinate of Q_n tends to the corresponding coordinate of Q:$x(Q_n) \to x(Q)$.

In the case of bounded precision, the neighbourhoods will themselves be of the above-mentioned form and we observe, in particular, that they are necessarily unbounded. In fact, they have cylindrical structure: if $\delta(Q)$ is the intersection of n half-spaces (as we can always assume, by virtue of what we noted above) and s is an arbitrary line contained in the intersection of the n hyperplanes which delimit the half-spaces (this intersection is always an infinite-dimensional space; one might say "all except n"), then together with any point P contained in $\delta(Q)$ are contained all lines through P parallel to s. To visualize this fact more easily, it suffices to note that two planes in S_3 cannot cut all the straight lines (more precisely, they do not cut those parallel to their line of intersection), and the same happens when one considers three hyperplanes in S_4, four in S_5, …, r in S_{r+1}: *a fortiori*, this happens for r hyperplanes in S_{r+2}, S_{r+3} and so on (and, in fact, the lines which are not cut will determine an infinite number of different directions; ∞^1, ∞^2 etc.). Finally, we note that this happens in any case where a space is cut by hyperplanes which number less than the dimension of the space: all the more so if the hyperplanes are finite in number and the dimension of the space is infinite.

9 Verifiability and 'Indeterminism'

The notion of *indeterminism* and the related notion of *complementarity* have arisen in the context of well-known 'anomalies' encountered in the study of physical phenomena. More specifically, in the study of phenomena where for certain aspects the particle interpretation is appropriate and for others the wave interpretation holds. Neither interpretation can lay claim to being universally acceptable, nor can the two be considered simultaneously without leading to 'contradictions'.

The question has been, and continues to be, a live topic of discussion; many-sided, and requiring special competence in several fields. The arguments put forward have offshoots in many directions, making it extremely difficult both to encompass them all (even if one restricts oneself to the essential points) and to single out with sufficient clarity either one topic, or a small group of them, on which one would like to concentrate attention.

The aspect which concerns us here is the logical-probabilistic one (and, in fact, for the time being, just the logical aspect, although with a view to the probabilistic side of things, for which it will serve as support). The study of this aspect could not be carried out, however, without touching upon points relating to other aspects and without indicating the position taken up with respect to them, a position which appears to correspond to that underlying the proposed choice of approach in the logical field.

Let us, without further ado, indicate which works we shall be referring to most frequently in what follows: on the one hand, that of von Neumann, and, in particular, the exposition and development given by Bodiou, whose formulation is in the field of direct interest to us; on the other hand, that of Reichenbach, who seems to me to present the questions most lucidly from the logical and philosophical point of view.[29] The solution

29 John von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press (1955) (a translation, by Robert T. Beyer, of *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin (1932)); J. von Neumann and G. Birkhoff, 'The logic of quantum mechanics' *Annals of Math*. (1937); G. Bodiou, *Théorie dialectique des probabilités* (etc.), Gauthiers-Villars, Paris (1964); Hans Reichenbach, *Philosophical Foundations of Quantum Mechanics*, University of California Press (1944).

References to these works in the present section will be indicated by means of initial and page number; R, p. 238, for example, for Reichenbach.

that I will put forward is a different one, but it could, in a certain sense, be seen as a simplified version of those set out by these authors.[30]

The different solutions, or interpretations, relating to the points we have to consider here, are concerned, explicitly, with the systematization in the logical domain of those kinds of statements which, because of their association with 'anomalies' like those mentioned above, lead to confusion. In order to incorporate them, it is suggested that, in general, we must have recourse to new logical structures, different from the usual structures, such as many-valued logics, or logics with modified operations and rules (in particular, 'nonmodular' logics).

The very starting points on which the analysis of these problems is based differ, however, one from the other. This difference is mainly between those who consider the problems as strictly peculiar to quantum physics, and who therefore pose the problems directly in terms of its technicalities, and those who see the problems as problems of thought in general. In the latter case, these problems could still appear more or less bound up with quantum physics, but only for contingent reasons; that is because they satisfy needs which actually arise in that theory (some would say 'exclusively' so, some 'mainly').

The formulation of von Neumann (vN, pp. 247–254) is strictly in terms of quantum theory, and takes as its starting point a Hilbert space (of functions ψ) in which the (linear Hermitian) operators correspond to quantities. An *event* is a quantity capable of assuming only the two values 0 and 1,[31] and therefore represented by a *projection-operator* E (which is idempotent, $E^2 = E$; that is having possible values – and eigenvalues – either 0 or 1); that is by a closed linear manifold $\mathscr{M}$ (that onto which E projects orthogonally). The event E is certain or impossible according to whether ψ belongs to $\mathscr{M}$ or is orthogonal to it; in all other cases, E has probability equal to the square of the projection of ψ onto $\mathscr{M}$. Two events are incompatible if they are orthogonal; they are simultaneously verifiable (not 'complementary') if they are commutative (in which case the logical product and logical sum are meaningful); and so on. To quote von Neumann (p. 253):

> 'As can be seen, the relation between the properties of a physical system on the one hand, and the projections on the other, makes possible a sort of logical calculus with these'.

---

30 My attitude had previously consisted in rejecting ad hoc interpretations in relation to quantum physics in order to reduce everything, essentially, to familiar situations (to facts which were 'complementary' in the sense that they were conditional on mutually exclusive experiments; like the behaviour of an object in two different destructive testing situations; or the victory of a tennis player in two different tournaments taking place at the same time in two different countries). A mention of this can perhaps only be found in the CIME (Centro Italiano Matematico Estivo) course given in Varenna, 1959. This solution seems to coincide with that of B.O. Koopman, *Quantum Theory and the Foundations of Probability* (1957).

Subsequent reflection (after a good deal of reading – the most relevant being that mentioned above), has not changed my original view, but rather made it more precise. In any case, it is, of course, simply an attempt at explanation (as we remarked at the beginning of the chapter) given the many issues involved, some of which may have escaped my notice.

31 Let me just mention, as an interesting curiosity, that this is the same convention as I had adopted (in a paper of 1964, and now here) after much hesitation, considering it novel and perhaps unacceptable. I subsequently realized that, far from being new, it had been in use since 1932 (together with all its developments). I wonder if the fact of its not being taken up confirms my doubts about its unacceptability?

The study of this kind of logical calculus (in terms of projections) has led (vN and B) to the identification of *nonmodularity*[32] as the characteristic property which distinguishes the lattice of this logic from that of standard (Boolean) logic.

A development which is inspired by the trend towards studying, in a more autonomous manner, or even completely separately, logic (and probability) on such a lattice – or on similar structures, also referred to many-valued logics – can be found in the work of Bodiou. His intentions are clearly summarized in the following passage (B. p.7):

> 'The primary motivation for our work, quantum theory, might appear contingent and particular, and capable of disappearing by the wayside if, by chance, quantum theory should come to be incorporated within a classical theory, which eliminates its "anomalies". This is what contemporary probabilists seem to believe and to expect. We shall attempt to show that they are wrong, and that the quantum calculus is simply a special case, imposed by necessity, of a general calculus of probability, which we call *dialectic*. This latter, far from being an unnatural growth on the body of the classical calculus, in fact subsumes it.'

Discussions of which statements and interpretations 'are or are not meaningful' are more directly considered, and more rigorously set out in Reichenbach, in a form which makes specific reference to quantum mechanics (and compares, in this context, the work of various authors), but which, from a conceptual point of view, can be adapted to any context whatsoever. For this reason, we shall develop our own analysis by using his (Reichenbach's) remarks as a guideline, putting forward our remarks as comments on his. In any case, the object of the analysis is that of finding the logical constructions that will prove suitable for resolving the difficulties in which we find ourselves; a topic which has attracted many currents of ideas from many different sources. This goal does not appear to have been achieved, nor does it appear that the efforts to reach it have opened up any promising avenues. I have the feeling that (as I said in my preliminary remarks in Section 6) the correct path is straightforward and simple, but it is my belief that it is obscured precisely by preconceived ideas about what it is that constitutes a necessary prerequisite for any logic.[33]

This would also appear to be a move in the direction of a natural continuation of a natural process; that of eliminating the drama from the initial state of confusion brought about by the appearance of something new, in contrast to one's accustomed way of seeing things. This has already happened for the Copernican system and for non-Euclidean geometry, for logical paradoxes and for relativity theory, and was bound to happen for the 'anomalies' of quantum physics. It seemed as if either *logic* itself was on trial or had

32 See, for example, L. Lombardo-Radice, *Istituzioni di algebra astratta*, Feltrinelli, Milan (1965), pp. 332 ff.

33 A comment seems called for at this point. My agreement or disagreement with the opinions of various authors concerns the individual points which necessarily arise in the course of an argument, and does not indicate any general position for or against. In every work there are inevitably a number of points with which the reader agrees or disagrees, either strongly or to some extent, or is in doubt about, or indifferent to, or simply does not understand. This also holds true for those works which I value to the extent of making them the basis for a discussion, an indication in itself of the stimulation I derived from them.

fallen apart completely. Reichenbach makes it clear, however, that logic, including probabilistic logic, is not to blame; on p. 102 he says:

> 'The rules of logic cannot be affected by physical experiences. If we express this idea in a less pretentious form, it means: If a contradiction arises in physical relations, we shall never consider it as due to formal logic, but as originating from wrong physical interpretations.'

The attributing of 'anomalies' not to the *structure of the physical world*, but rather to the 'structure of the languages in which this world can be discussed' is even more decisive. 'Such analysis expresses the structure of the world indirectly but in a more precise way' (R, p. 177). These are the languages which, by means of *definitions*, introduce into the world of observable phenomena something that we might call 'interphenomena' (non-observables). As examples of such 'definitions', consider those which attach to the 'observed value' the meaning of being the value of the quantity *before and after*, or only *after*, and so on. As examples of such languages, consider the particle language, the wave language and a neutral language: The first two '… show a deficiency so far as they include statements of causal anomalies, which … can be transformed away, for every physical problem, by choosing the suitable one of the two languages. The neutral language is neither a corpuscle language nor a wave language, and thus does not include statements expressing causal anomalies. The deficiency reappears here, however, through the fact that the neutral language is three-valued; statements about interphenomena obtain the truth-value *indeterminate*' (R, p. 177). The same situation is described by Bodiou as the existence of several 'coherent formal systems', like the mechanics of points, and the wave theory of light. Two 'attributes' pertaining to different systems, like a statement in particle form and one in wave form, 'might be *incoherent* without being *contradictory*' (B, p. 11).

The appearance of the word 'indeterminate', or of the distinction between 'incoherent' and 'contradictory', indicates that in order to find something suitable for our purpose we must bring into the world a new logic. There is no difference, in principle, between the approaches of the two authors cited. The one introduces straightaway a third 'logical value' and then goes on to define the logical operations by means of 'truth tables'; the other defines the operations axiomatically and could (it seems to me – in fact, he does not, although I think he should do so) define the 'truth values' on the basis of them.

Reichenbach distinguishes two variants depending on whether a statement that is neither *true* nor *false* is called *indeterminate* or *meaningless*. The different names do not correspond to different meanings of the partitions into the three cases; the change in name corresponds to the case in which 'it is necessary to make an observation H in order to know whether E is true or false'. One agrees then to say that

E is *true* if observation H has given the result E,
E is *false* if observation H has given the result not-E

$$E \text{ is } \left\{ \begin{matrix} \textit{meaningless} \\ \text{or } \textit{indeterminate} \end{matrix} \right\} \text{ if the observation } H \text{ has not been made.}$$

Using the notation introduced for conditional events, this turns out to be exactly what we agreed to say by writing $E|H$ instead of E, and putting $E|H = 1$ (true), or $= 0$ (false), or $= \varnothing$ (void) (and, if we wish, we could call it 'meaningless' or 'indeterminate' instead of 'void').

The meaning of the trichotomy does not depend at all on which words we use; the way in which it is defined is the only thing that matters. It might be conceded, however, that it does make some difference whether we use 'meaningless' instead of 'indeterminate'. It is a difference of philosophical attitude – an acceptance of the Bohr or Heisenberg interpretation. And there are formal consequences if one believes that a meaningless statement cannot even be mentioned, whereas by calling it *indeterminate*, and considering this, as '*an intermediate truth-value*' (R, p. 145) lying between true and false, it becomes permissible to speak of it and *above all to work with it*.

This is, in fact, the requirement that must be satisfied if something is to be called a mathematical structure or, in particular, a logical structure. It is very easy to construct such a structure. Considering the tableau on the left, there are 3^9 (=19 683) different ways of substituting the letters T, F, I (True, False, Indeterminate) in place of the asterisks and one can choose a subset of these to which to assign the title of operation and a symbol (e.g. +) to replace the symbol ° between A and B at the corner.

$A \circ B$	B: T	F	I
A: T	*	*	*
F	*	*	*
I	*	*	*

$A + B$	B: T	F	I'
A: T	T	T	T
F	T	F	I
I	T	I	I

The table headed $A + B$ is thus filled with entries that are either T or F or I, placed in the 1st, 2nd, or 3rd row according to whether A is true, false, or indeterminate, and in the 1st, 2nd or 3rd column depending on the value of B (an example is given in the tableau on the right). Reichenbach (p. 151) introduces seven such binary operations (some of which are taken from the work of Post): disjunction and conjunction (extensions of logical sums and products), three forms of implication (the standard one, an alternative and a form of quasi-implication), two forms of equivalence (the standard one and an alternative), and three unary operations of negation (cyclical, diametrical and complete). Four of these operations are due to Reichenbach himself, but he leaves out 'some further implications' defined by Post. Variants due to other authors are also mentioned.

Without going into a more detailed discussion (which would lead on to more substantial objections), we note that all this could be expressed in terms of two-valued logic by thinking of a 'three-valued event' (in our terminology a 'conditional event', but nothing is altered if one refers to it – or thinks of it? – in a different way), E, say, expressed in the form $E'|E''$, or in that of the partition into three cases in which it is either true, false or void:

$$E^T = (E = 1) = E'E'', \quad E^F = (E = 0) = \tilde{E}'E'', \quad E^I = (E = \varnothing) = \tilde{E}''.$$

An E whose logical value depends on the logical values of two other 'three-valued' events, A and B, is obtained by defining as logical functions of A^T, A^F, A^I, B^T, B^F, B^I, both E' and E'', and E^T, E^F, E^I (should they turn out to be exhaustive and exclusive), and putting $E = E'|E''$ or $E = E^T|(E^T + E^F)$.[34] In this way, one avoids creating a number of symbols and names of operations and consequent rules (which are difficult to remember and sort out, and difficult to use without confusion arising). Above all, one avoids creating the tiresome and misleading impression that one is dealing with mysterious concepts which transcend ordinary logic.

Bodiou (like, of course, many previous authors whose approach and notation he follows) does not base his work on three truth-values (although, in his notation, the three possibilities for a proposition a seem to me to be expressible in the form a, Ca, $C(a \vee Ca)$, where C denotes negation). He does not even have a symbol for the 'third value' (whereas he uses u, True, and $\varnothing$, False, corresponding to our use of 1 and 0). This takes one even further from an immediate understanding of the meaning. There is also a section (B, pp. 30 ff) devoted to many-valued logics in which there are M 'truth-values', which can be denoted by $k/(M-1)$, $k = 0, 1, 2,\ldots, M-1$, but not even here is there a value equivalent to 'Indeterminate'. The work, in fact, proceeds in an entirely different direction, in which the value $W(a)$ of a proposition should have a meaning something 'similar' to 'probability' (but with its own 'rules': $W(a \vee b) = W(a) \vee W(b)$, with the same for $\bigwedge$; fortunately, we have $W(Ca) = 1 - W(a)$).[35]

With this, the time has now come (for two reasons which are formally identical, but as far as interpretation goes are totally unrelated) to examine the real meaning of these questions; no longer just formally, but in depth. And it is at this point that, in our examination, we must take into consideration, together with the notion of indeterminacy, the notion of *complementarity*.

10 Verifiability and 'Complementarity'

The essential problem, the basic doubt which came to the surface in the previous analysis, can be expressed formally in the following way. Suppose we have two or more 'three-valued events' (we shall consider them as conditional events, but it does not make any difference), and let us denote them by $E_1 = E_1' \mid E_1''$, $E_2 = E_2' \mid E_2'', \ldots, E_s = E_s' \mid E_s''$. Can it be meaningful and interesting to define another such, an $E = E'|E''$, whose meaning is related to the meaning of the others? And, in this case, will its 'truth-value' be a function of those of the others?

If we think of the general case (for example, of s conditional bets) it is likely that a few events (simple, two-valued ones) will be of interest; like $E_1^T E_2^T \ldots E_s^T$, $E_1^T + E_2^T \ldots + E_s^T$,

34 Note that E'' is the same thing as $E^T + E^F$, whereas for E' it does not matter whether we take E^T or $E^T + E^I$, or any event in between ($E' = E^T + D$, with D contained in E^I).

35 Such rules, proposed by other authors, are changed by Bodiou in a way which brings them closer to probability theory (but then, of course, *one no longer has operations on logical values*).

and the similar forms with E_h^F or E_h^I (expressed by means of simple events), which express the fact that either all the events, or at least one, are won, or are lost, or are called-off (because the conditioning event did not occur).

If we think of cases of actual interest, however, it appears that we have to reconsider our whole approach to the problem, since some Es may arise that are connected with the E_h in a meaningful way, although not necessarily with their logical values. As a trivial example – but, for this very reason, instructive – let us begin by considering the (two-valued) events E_h', and for each of them construct the (two-valued) event E_h'' = 'I know (at this moment) whether E_h' is true or false'. By this means, we have transformed our field of events into a field of three-valued events, in which the third value stands for 'unknown', whereas True and False stand for 'known as true' and 'known as false'. One often emphasizes – and for good reason – that 'Indeterminate' is not to be confused with 'unknown' (see, e.g., R, p. 142). The same thing can be underlined in a rather better way by saying that the two notions coincide only in this particular example. One might say that this example corresponds to thinking of what 'I know at this moment' as frozen (I will no longer be able to learn anything about that which is now unknown to me, and I have no further interest in it; for me it will be for ever indeterminate, or even 'meaningless').

In this example it is natural to call the logical sum of two conditional events E_1 and E_2, $E = E_1 \vee E_2$, the conditional event corresponding to the logical sum $E_1' \vee E_2'$ of the corresponding events E_1' and E_2'. We have, therefore, $E = E'|E'' = (E_1' \vee E_2')|E''$, where E'' = 'I know (at this moment) whether $E_1' \vee E_2'$ is true or false' (i.e. if at least one of E_1' and E_2' is true), and this does not coincide with $E_1'' \vee E_2''$ (although it necessarily contains it) since one might well know, for instance, that someone 'arrived yesterday or today', but not know which. It follows, therefore, that the event $E = E_1 \vee E_2$ thus defined is *not* a logical function of E_1 and E_2 in the sense we have seen so far (function of their logical values). E is certainly true if at least one of the E_h is (certainly) true, certainly false if they are both known to be false; but if they are both indeterminate (unknown) it could be either indeterminate or (known to be) true. (And if there are more than two of the E_h, one has the latter case if at least two are indeterminate and the others false.)

This example, as we have said, is trivial; but the cases in which considerations of this kind find an actual important application are precisely (I would even say exclusively) those modelled on the same scheme (except that 'I know …' is replaced by 'it has been verified that …', or 'it will be verified that …' – within a certain time period, for example – and so on).

Quantum theory provides an obvious example of a case in which everything is more clear-cut. If E_1 and E_2 denote two events (or, equivalently, the respective projection-operators), an observation by means of the operator $E_1 \vee E_2 = E_1 + E_2 - E_1E_2$ is an observation for the event-sum, but not for either of them individually. Similarly, one can make an observation of $X + Y$ or XY and so on without making observations of the two separate random quantities X and Y. The concept is an analogous one but, if we wish to confine ourselves to projection-operators representing events, we must restrict ourselves to considering the events consisting of whether or not X (or Y, or $X + Y$, or Y etc.) belongs to a given interval, $E = (a \leqslant X \leqslant b)$; that is $E = 1$ if X lies between a and b, and $E = 0$ otherwise.

When we turn to verifiability (in the various senses considered in previous sections) the situation is similar. Leaving aside the details and the finer points (we do not have to repeat them here, having dwelt upon them – perhaps at too great a length – already), it will suffice to refer to the 'trivial example' considered above, taking as 'indeterminate' that which will be 'known (not now, but) after a certain time, or after certain checks have been made, or after having obtained some given information, and so on.' Here, too, '*E* is indeterminate' is the statement of an objective fact; the lack of the information required in order to decide the truth or falsity of *E* within the specified time period and according to the rules laid down for doing so. The difference is that the assignment of the value 'indeterminate' (and the same for 'true' and 'false') is, in this sense, not immediate (as it was in the trivial example): it is not excluded, however, as in two-valued logic (where 'unknown' is always considered as a temporary state of affairs, pending knowledge of the truth – even if the period of waiting should turn out to be in vain, or to last for ever).

Within this context, it appears to be possible to pose the problem of complementarity and to discuss its real meaning. Let us first of all observe that, in the sense we have just used it, the qualification 'indeterminate' might be attributable to an event *as of now*: that is when, on the basis of what we already know about the events, and about the possibilities and known means of obtaining information, we are in a position to exclude the possibility of getting to know whether *E* is true or false within the given time period, and according to the specified mode of doing so. Clearly, only a few of the 'unknown' events (or *possible* events, as we used to call them) will be 'as-of-now indeterminate' (in exceptional cases all of them might be and this would reduce to the trivial example considered above). We could say, in order to be more precise, that the division (at a given moment) of the events into certain, impossible and possible (i.e. with values known, as of now, to be true, false or unknown) could be pursued further, subdividing the possible (unknown) events into five subcases depending on the situation considered as 'final'. Specifically, we can distinguish the events which will eventually be certainly indeterminate (the case *I* already mentioned), or those for which there is doubt between the outcomes *T–I* (true–indeterminate), *I–F* (indeterminate–false), *T–F* (true–false), *T–I–F* (true–indeterminate–false).

From these conclusions concerning single events, we can pass to the properties of two or more possible events. When we were restricting ourselves to True and False, for example, we were able to say whether two or more events were incompatible or exhaustive. This meant that although it was possible for any of the events to occur, not more than one of them actually could. In the same way, it is possible that in the case of indeterminacy similar exclusions can be made. It may be certain, as of now, that, from among two or more events, each of which might or might not in the end turn out to be indeterminate (i.e. they are all either *T–I*, *I–F* or *T–I–F*; none of them are *I* or *T–F*), at least one remains indeterminate, or at least one does not, and so on, and so forth.

The interesting case in practice is that of two events, one at least of which remains certainly indeterminate (but it is not known which; otherwise it is easy to couple it with another one). Two such events are called *complementary* (and a similar definition holds for quantities, as we shall shortly see). The purpose of the more general discussion given above was merely to show how the notion corresponds to a natural examination of the possibilities that present themselves when we extend the classification by introducing 'indeterminate' as a third logical value.

Complementary events arise, for example, when establishing:

whether or not a tennis player wins if he takes part in one or other of two tournaments taking place at the same time in two different countries;

whether a coin will show Heads or Tails the next time it is tossed, assuming that the next toss is performed by either Peter or John;

what are, fixing them in one's mind, the registration number and the features of the driver of a suspect car that flashes past (assuming that it is at best possible to observe one or other of the two items);

what is the behaviour of one and the same object when it is subjected to one or other of two destructive tests;

whether a given building (e.g. the Tower of Pisa) will remain standing until some specified date under the assumption that some kind of repair work is carried out (or assuming some other project); and so on.

As well as events, one could equally well speak of *complementary quantities*. Referring to the above examples, we could consider the remaining life of the Tower of Pisa conditional on one or other of the hypotheses considered (of which only one will be observable – that conditional on the course of action actually chosen). Two random quantities X and Y are, by definition, noncomplementary (i.e. simultaneously measurable) in the strict sense, if this condition holds for all the events $X \leqslant x$, $Y \leqslant y$, for arbitrary x, y (and, in quantum theory, this reduces to the spectral decompositions of the corresponding operators; see vN, Chapter II, and, for noncomplementarity, p. 254). An amusing example, but one which well conveys the idea (in a nutshell), is the complementarity of the two measurements that a tailor would have to make simultaneously when one of them requires the client to hold his arm straight downwards, and the other requires that he hold it parallel to the floor and with the elbow bent to give a right angle.

The most celebrated example is undoubtedly that of complementarity in quantum mechanics, and there is no doubt that this is the most important case, because of the profound nature of the implications regarding our conception of the nature of phenomena and the knowledge of them that we can attain. There is also a more 'technical' and precise way of expressing the condition of complementarity for events in this case. As we have already mentioned, E_1 and E_2 are noncomplementary events if, as projection-operators, they commute. Does this (together with related factors) provide sufficient justification for the idea that one has to make a *distinction of a logical nature* between complementarity in the realm of everyday affairs and in that of quantum physics? The answer would seem to be no. Otherwise, why should we not say that incompatibility – corresponding as it does in the quantum theory formulation to orthogonality – should be considered in that context as something completely different, even though it is exactly the same thing? This comment is certainly not sufficient to settle the argument; nor is the much more basic fact that we have up to now presented the notion of complementarity without having encountered any need to introduce such distinctions. We shall have to be more specific, albeit in a summary fashion, about the physical meaning of the problems under consideration, while, on the other hand, we must examine, from a critical standpoint, the arguments put forward on the basis of these physical considerations to support the opposite point of view.

As a first step, we shall simply develop the *description* of what we shall need as background for our purposes. This can then be used as a basis for the understanding of the logical situation and also by those who are not familiar with the physical and mathematical interpretations of the schemes on which we shall base our considerations. We shall take as our starting point the remarks of Section 9, concerning the interpretation of events as projection-operators and, in fact, we shall restate this, for the convenience of the reader, and provide an integrated and extended version by including the case of quantities. Notwithstanding the inevitable fact that many things will be glossed over, this should be sufficient, and the resulting picture should turn out to be clear and precise enough for the purposes for which it is intended.

11 Some Notions Required for a Study of the Quantum Theory Case

As the fundamental notion, we take a space $\mathscr{H}$ which, by means of its points, or rather vectors, provides a suitable representation of the 'states' in which a given physical system S can find itself. The space is like ordinary three-dimensional space, with all the affine and metric properties (those of analytic geometry) but is infinite-dimensional (*Hilbert* space). Its points, or vectors, represent functions (functions defined on the space of possible configurations of the system; for example, of three coordinates x, y, z in the case of a single free particle, and of $3N$ in the case of N distinct particles). The 'state' of the system, at a given instant, is characterized by one of these functions, its ψ-function (or ψ-point, or ψ-vector, as we call the point, or vector, which represents it in the space $\mathscr{H}$); ψ is such that the vector has modulus $\|\psi\| = 1$.

The way the system evolves in time is described by the variation of ψ as a function of time (deterministically set out by equations similar to those of classical mechanics). The difference is that these equations no longer tell one how the configuration of S varies, as it is and as it is observed, but only how the probability of finding it in this or that configuration varies if one submits it to an 'observation' at some future time t.[36]

36 Let us just mention some of the omitted details. The functions ψ (and, in general, all the functions considered) are complex (roughly speaking, for the same reason as it is convenient to express oscillations in terms of $e^{i\omega t}$ rather than $\cos \omega t$), and, as such, considered as vectors, they have bounded moduli ($f^2 = \int |f|^2 \, dS < \infty$). The space of these functions is the *Hilbert space* with the *Hermite* inner-product ($f \times g = \int f.g^* \, dS$, where the asterisk denotes the complex conjugate; we always have $(f \times g) = (g \times f)^*$ and $|f \times g| \leqslant \|f\| \, \|g\|$). One can directly characterize an infinite-dimensional linear metric space as a *Hilbert* space by adding the properties of completeness and separability. With a system of (orthogonal, etc.) Cartesian coordinates, it is the space of points defined by sequences of coordinates $x_h (h = 1, 2, \ldots, n, \ldots)$ such that $\sum_h |x|^2 < \infty$ (and this expression gives the modulus of the vector with coordinates x_h; the Hermitian inner-product of two vectors is given by $\sum_h x_h y_h^*$). The linear operators that we shall come across are also Hermitian (or self-adjoint: $A^* = A$, where A^*, the adjoint of A, is defined by $A^* f \times g = f \times Ag$); the operators can be represented by matrices (with reference to an orthogonal Cartesian system) with entries A_{rs} (and then $(A^*)_{rs} = (A_{sr})^*$; A is Hermitian if $A_{sr} = (A_{rs})^*$). See vN, especially pp. 34–46, and for a more direct exposition and interpretation, E. Persico, *I fondamentali della meccanica atomica*, Zanichelli, Bologna (1936).

Let the above serve to give an idea of the various detailed specifications which, like the present one, would be out of place in the main text if they were to give the reader the impression that he has to acquire a knowledge of these notions, or to refresh his rusty memory, or to worry about the details, in order to understand those few points on which his attention would be better focused. And let it serve also as a warning for those who were tempted to accept the present formulations literally, or to be put off at finding them incomplete.

We cannot 'see' the vector ψ; it is a mathematical abstraction in a space that is a mental fiction. But, by starting with the results of the last observations made (and assuming that the system S has not, in the meantime, been subjected to any external disturbance), and knowing the laws governing its evolution, we can, in principle, determine it. In any case, assuming that we knew the vector ψ, little or nothing would be known about the actual configuration of the physical system S in which we are interested, and about its evolution.

More precisely – in order to make clear what ψ is or is not sufficient to explain – let us consider an arbitrary event E, that is any statement whatsoever concerning the space S at a given instant (this must always be understood, even if not mentioned explicitly; two events or two quantities of the same kind, but relative to different times, are two distinct events or quantities). What we can say about E is that it is certainly *true* if the vector ψ belongs to some given linear manifold $\mathscr{M}$ (associated with E), and certainly *false* if it belongs to the linear space of all vectors orthogonal to M. In these two cases it is superfluous to make an observation, because the answer would certainly be the one we have just given (but it would also be innocuous because it would not disturb the system at all). If, on the other hand, the vector ψ is neither contained in nor orthogonal to the space $\mathscr{M}$ an observation concerning E gives an unforeseeable result. Knowledge of the state, which is all contained[37] in the vector ψ, does not determine this result in advance, but it is not without value, because it gives all that can be given: that is, the *probability*. The details are as follows: we decompose the vector ψ into two components, one parallel to $\mathscr{M}$, the other orthogonal; that is into $\psi = E\psi + \tilde{E}\psi$ (in this way indicating the projection E onto $\mathscr{M}$, and the orthogonal projection, $\tilde{E} = 1 - E$, onto $\mathrm{H} - \mathscr{M}$). The probabilities of the E and $\tilde{E}$ that result are given by the squares of the respective projections:

$$\mathbf{P}(E) = E\psi^2, \quad \mathbf{P}(\tilde{E}) = \tilde{E}\psi^2.$$

Note that, instead of $||E\psi||^2 = E\psi \times E\psi$ (1st form) we can also write $E\psi \times \psi$ (2nd form, which equals $E\psi \times (E\psi + \tilde{E}\psi)$, whereas $E\psi \times \tilde{E}\psi = 0$), which is also valid even if E is not a projection-operator and has a similar meaning even in this case (as we shall see).[38] In the case where E is a projection-operator, one verifies immediately that $\mathbf{P}(E) + \mathbf{P}(\tilde{E}) = 1$, as was necessary. In fact (because of the orthogonality of the two components, and hence by Pythagoras), we have

$$\mathbf{P}(E) + \mathbf{P}(\tilde{E}) = E\psi^2 + \tilde{E}\psi^2 = \psi^2 = 1.$$

The interpretation of E as a projection-operator turns out to be even more expressive, however, in light of the following. The most important fact is that the vector ψ is not restricted to a passive rôle of indicating the probability of the required result being observed. The observation itself is forced to choose whether to fall into the space $\mathscr{M}$

37 See the comments that are made later following the discussion of the possibility of explanations introducing 'hidden parameters'.

38 The equivalence between the two forms does not hold in the general case. There, in fact (see footnote 36), we have $A\psi \times A\psi = A^*A\psi \times \psi$, and, in the case which interests us (the Hermitian form), this equals $A\psi \times \psi$. In order that it equals $A\psi \times \psi$, we must have A idempotent; i.e. A = projection-operator (with all the eigenvalues idempotent, $\lambda = \lambda$; that is, $\lambda = 1$ or $\lambda = 0$).

corresponding to E, or into the orthogonal space $\mathscr{H} - \mathscr{M}$ corresponding to $\tilde{E}$ (and it does so with the probabilities indicated). The outcome simply constitutes the information about the choice made. From the position taken up after the jump we have obliged it to take, the system returns to an evolution according to the previous laws until there is a new disturbance.

This picture (anthropomorphic, but this perhaps helps one grasp the ideas in the absence of a more detailed technical exposition) contains 'in a nutshell' all that is required for a complete treatment. It will suffice, essentially, to consider simultaneous questions about several events, rather than a single one (this will also hold for measurements of quantities), and to distinguish the cases which give rise to the circumstances to be discussed.

Instead of a partition into two opposite events (E and $\tilde{E}$) we can think of a 'finer' partition, still 'complete', into some number n of incompatible events, $E_1, E_2, \ldots, E_n$, or even into an infinite number. Each E_h will be defined by the corresponding (closed) linear space $\mathscr{M}_h$, and all these spaces must be taken to be orthogonal to each other, and such that taken altogether they form the entire space $\mathscr{H}$ (i.e. there must not exist a vector in H orthogonal to all the $\mathscr{M}_h$, and then there will not exist vectors which are not linearly dependent on the vectors of the $\mathscr{M}_h$). The total dimension is countable, being coincident with that of $\mathscr{H}$. It follows that, in the case of a finite partition, at least one of the $\mathscr{M}_h$ must be infinite-dimensional (in particular, note that for E and $\tilde{E}$ either one is infinite-dimensional or both are). In the case of a countable partition, this is not necessarily the case, and we can even consider the extreme case where all the spaces $\mathscr{M}_h$ are one-dimensional.

This is the fundamental case in terms of which the discussion of all the others is framed. In other words, it is the case in which we have a system of orthogonal Cartesian axes corresponding to an infinite set of events $E_h (h = 1, 2, \ldots, n, \ldots)$, interpretable, for example, as the (distinct) values, λ_h, which a quantity Z can assume: $E_h = (Z = \lambda_h)$. On the other hand, we clearly have $Z = \sum_h \lambda_h E_h$ (as a random quantity), since one and only one of the E_h will occur (and will take the value 1; all the others will be 0), and the sum will reduce to the corresponding value λ_h. Defining Z as an operator (associated with the quantity of the same name) in the same way, it seems clear from the identity of the written forms that one is dealing with the operator formed by multiplying the axis vectors (functions) E_h (the eigenvectors or eigenfunctions) by the λ_h (the eigenvalues). This gives the prevision (or mathematical expectation) of the quantity Z by means of the same formula used for E (2nd form):

$$Z\psi \times \psi = \left(\sum_h \lambda_h E_h\right)\psi \times \psi = \sum_h \lambda_h \left(E_h \psi \times \psi\right) = \sum_h \lambda_h \mathbf{P}\left(E_h\right) = \mathbf{P}(Z).$$

In a similar fashion, one obtains, immediately, the distribution function of Z: putting $E_z(\lambda) = \sum E_h(\lambda_h \leqslant \lambda)$, one has the event $E_z(\lambda) = (Z \leqslant \lambda)$, or the related projection-operator; it follows that

$$E_z(\lambda)\psi \times \psi = \mathbf{P}\left(E_z(\lambda)\right) = \mathbf{P}(Z \leqslant \lambda) = F_z(\lambda),$$

where we denote by F_z the distribution function of Z.

The collection of projection-operators $E_z(\lambda)$ (or the set of their linear spaces, each of which, if we proceed in the direction of increasing λ, contains all the preceding ones) defines the spectrum of Z; in this case, a discrete spectrum (there are a countable number of values of λ_h).

Going back to the physical problem, we can repeat what was said in the case of an event. If ψ belongs to one of the axes, and only then, the corresponding event E_h will certainly be true; that is, thinking of the quantity Z, its value will be λ_h with certainty. It is unnecessary (but harmless) to make an observation. In all other cases, the vector ψ will be forced (if we make an observation on all the E_h together, i.e. on Z) to choose along which axis it is to lie: the result (E_h, or λ_h) indicates which choice was made, and, after the jump, we go back to the normal evolution.

There is one difference, and a very important one, with respect to the previous case. Now we know exactly,[39] after the observation, the position chosen by ψ (whereas, beforehand, we knew only that it belonged to $\mathscr{M}$, or alternatively to $\mathscr{H} - \mathscr{M}$). This happens only in the case now under consideration; that of a partition which gives rise to spaces which are all one-dimensional. When one defines on them a quantity Z, it is necessary that the λ_h values (the eigenvalues) are all distinct (simple), otherwise, the refinement of the subdivision we have reached will in part be destroyed (one has the case of 'degeneracy'). For the case we are dealing with (the nondegenerate case) one says – for obvious reasons – that a 'maximal observation' has been made.

We now turn to the problem of complementarity of observations.

Can we ask for the simultaneous verification – that is with one and the same observation – of two or more events? Or for the measurements of two or more quantities? (And we note that 'simultaneously' can only mean 'with one and the same observation'; another observation made immediately afterwards would already find ψ changed by the effect of the first one.)

The answer is obvious if one thinks of two 'maximal observations', like the measurements of two quantities Z' and Z'', to be performed simultaneously. In doing so, we force ψ to lie on one of the axes of the first system and also on one of the second system. Now ψ obeys any order whatsoever, but cannot accept contradictory orders – and this would be the case if the two systems of axes do not coincide. In such a case, Z' and Z'' cannot be measured simultaneously; that is they are *complementary*.[40] If the axes do coincide, the result is trivial, because Z' and Z'' are functions one of the other (if Z' assumes the value λ'_h, it means that the hth-axis has been chosen, and hence that Z'' takes on the value λ''_h). The coincidence of the system of axes implies commutativity (in terms of operators, the condition is $Z'Z'' = Z''Z'$), and the same also holds in the case of events, $E'E'' = E''E'$, or of non-maximal quantities, $XY = YZ$. Non-maximal quantities X, Y relating to one and the same system of axes (suppose it to be that of Z) are obtained by taking eigenvalues μ_h and ν_h, which are not all distinct (so that X and Y as functions of Z, $X = f(Z)$, $Y = g(Z)$, are not invertible). Conversely, if X and Y are not complementary, and have as possible values the μ_i and ν_i, respectively, one can construct a Z (corresponding to a maximal observation) of which X and Y are functions, and having distinct values λ_h corresponding to all the compatible pairs (μ_i, ν_j). In particular, in the case of events, noncomplementarity, $E'E'' = E''E'$, means that the corresponding spaces $\mathscr{M}'$ and $\mathscr{M}''$ are mutually orthogonal (i.e. if we call M the intersection, $\mathscr{M} = \mathscr{M}' \mathscr{M}''$, then two vectors,

39 The ψ is, in fact, uniquely determined, because a multiplicative constant (real or complex) is irrelevant.

40 Think of the example of the tailor: complementary measurements are those that to be made simultaneously would require the client to simultaneously assume several different, incompatible positions!

one from $\mathscr{M}' - \mathscr{M}$, and one from $\mathscr{M}'' - \mathscr{M}$ – i.e. from $\mathscr{M}'$ and $\mathscr{M}''$, respectively – and orthogonal to M, are always orthogonal to each other; then, in fact, and only then, is the product of the two projections the projection onto the intersection M, and does not depend on the order). As special cases, one has the case of inclusion (if $\mathscr{M}' - \mathrm{M} = \{0\}$, we have $\mathscr{M}' \subset \mathscr{M}''$, $E' \subset E''$), and that of incompatibility ($\mathscr{M} = \{0\}$, $E'E'' = 0$ = impossible).

Noncomplementarity between X and Y can be interpreted in the same way because it is equivalent to the noncomplementarity of each of the events (projection-operators) $E_x(\mu)$ and $E_y(\nu)$; in other words,

$$E_x(\mu)E_y(\nu) - E_y(\nu)E_x(\mu) = 0$$

for any μ and ν whatsoever. Geometrically, this is equivalent to the orthogonality of the spaces $\mathscr{M}'_\mu$ and $\mathscr{M}''_\nu$ (i.e. orthogonality between the vectors of $\mathscr{M}'_\mu - \mathscr{M}'_\mu \mathscr{M}''_\mu$ and of $\mathscr{M}''_\nu - \mathscr{M}'_\mu \mathscr{M}''_\nu$). One could consider the same condition in a weaker version (limiting ourselves to checking the validity for certain values of μ and ν instead of for all of them), but we shall consider this in the context of 'continuous spectra', where it is more interesting.

The case of the 'continuous spectrum' arises with a quantity that can assume any value (between $-\infty$ and $+\infty$, or in an interval etc.), rather than just a finite or countable set of values as considered so far. In quantum physics one deals with nonquantized quantities (like the coordinates) in addition to the quantized ones (like energy).

In this case, too, in considering quantities X, Y,..., everything can be expressed by $\mathrm{E}_x(\mu)$, $\mathrm{E}_y(\nu)$,..., except that an $\mathrm{E}_x(\mu)$ will actually vary for all increments in μ (and not just in going through certain values of μ; the eigenvalues $\mu = \mu_i$) and the distribution function

$$F_x(\mu) = \mathbf{P}(X \leqslant \mu) = E_x(\mu)\psi \times \psi$$

will, in general, turn out to be continuous.[41] A decomposition into a finite or infinite number of incompatible events could be obtained by dividing the axes of the μ in some fashion into intervals $\mu_i < \mu < \mu_{i+1}$ ($i = 0, \pm 1, \pm 2, \ldots$, in order to denote them in increasing order, and letting them be, in general, unbounded in both directions). Only in this way can we have a partition

$$E_i = (\mu_i < X \leqslant \mu_{i+1})$$

that gives us, in the way we indicated previously, an 'approximate measurement' $\hat{X}$ of X, defined by choosing a subdivision μ_i, and in each interval a value $\hat{x}_i$, and then setting $\hat{X} = \sum_i \hat{x}_i E_i$. In other words; $\hat{X}$ is the function of X defined by the step-function

$$f(X) = \sum_i \hat{x}_i (\mu_i < X \leqslant \mu_{i+1}).$$

From this point of view, X is not a measurement with absolute precision, but with arbitrarily high precision if we substitute for it an $\hat{X}$ defined by a function with arbitrarily small steps. An observation on $\hat{X}$ forces ψ to lie in one of the spaces $\mathscr{M}_i$ (corresponding to $E_i = E_x(\mu_{i+1}) - E_x(\mu_i)$) and it is never maximal because the subdivision could always be made finer.

41 We ignore the mixed case of probability in part concentrated, in part continuous, etc. (see Chapter 6, 6.2.2–6.2.3).

If we now consider X and Y (both with continuous spectra), the condition for noncomplementarity is still the commutativity of X and Y as operators, $XY - YX = 0$; in other words, commutativity between the $E_x(\mu)$ and the $E_y(\nu)$. If the condition holds, then, apart from obvious complications, what was said in the case of the discrete spectrum also applies here: for example, it is also true in this case that one can construct a Z of which X and Y are functions, but that one can only obtain it (clearly) by means of procedures based on the Peano curve or something similar (vN, p. 178).

If, however, we content ourselves with approximate measurements, $\hat{X}$ and $\hat{Y}$, then the orthogonality of $E_x(\mu_i)$ and $E_y(\nu_j)$, relative to the points of subdivision chosen for the μ and ν, is sufficient for noncomplementarity; this weaker condition may also hold if X and Y are complementary. In other words, complementarity does not necessarily exclude the possibility of simultaneous approximate measurements; that is of two suitably chosen $\hat{X}$ and $\hat{Y}$.

And at this point we arrive at the special case of quantum mechanics, where complementarity often arises in the particular guise of noncommutativity, expressed by

$$XY - YX = h/2\pi i \quad \left(h = \text{Planck's constant}\right).$$

This holds where X and Y are coordinates, and for a conjugate impulse, or, more generally, in the terminology of classical mechanics, for 'canonically conjugate' quantities.

From this relation of noncommutativity (and hence complementarity between X and Y) we can derive a justification of Heisenberg's Uncertainty Principle, which indicates the way in which the precision of the measurements of X and Y – which can be made arbitrarily high if performed separately – turn out to have a reciprocal relationship under a simultaneous observation.[42]

The following is just a brief development of this crucial point. From $XY - YX = a$, it follows that

$$\left(XY - YX\right)\psi \times \psi = a\psi \times \psi = a\psi^2 = a \ \left(\text{if}\,\psi = 1\right).$$

We note also that

$$XY\psi \times \psi = Y\psi \times X\psi \leqslant Y\psi \cdot X\psi,$$

with a similar result for YX. It follows (by the triangle inequality!) that the bound for the difference is given by

$$\left(XY - YX\right)\psi \times \psi \leqslant 2.\left\|Y\psi\right\| \cdot \left\|X\psi\right\|,$$

that is

$$\left\|Y\psi\right\| \cdot \left\|X\psi\right\| \geqslant \frac{1}{2}\left|a\right| = \frac{h}{4\pi}.$$

42 The procedure which is briefly outlined here is taken from vN (p. 230, ff.), and is there (note 131, p. 233) attributed to ideas of Bohr, and work of Kennard and Robertson.

The inequality holds no matter where we take the origin for X and Y, and, in particular, if we take it at the mean value. The two moduli then have an interpretation as standard deviations, and we have the uncertainty principle in its usual form: $\sigma_x\sigma_y \geqslant h/4\pi$.

This means that it is impossible to approximate X and Y by means of some $\hat{X}$ and $\hat{Y}$ by choosing arbitrarily small 'steps' for both of them: their order of magnitude must be such that the product (as an order of magnitude) is not less than h. Geometrically, the subdivision into rectangles in the (X, Y)-plane, a consequence of the subdivision by means of the μ_i and ν_j, respectively, on the x-, y-axes, cannot be so fine as to give rectangles whose areas have orders of magnitude less than h (the choice of the ratio of height to width remaining arbitrary). These rectangles are the regions for which it can be 'verified' whether the pair of measurements fall inside or not: $(\hat{X} = \hat{x}_i)(\hat{Y} = \hat{y}_j)$ is, in fact, equivalent to $(\mu_i < X \leqslant \mu_{i+1})(\nu_j < Y \leqslant \nu_{j+1})$. Observe that this is precisely one of the conditions of 'bounded precision' considered in Section 8 (Figure A.1).

On the basis of the digression, which we are now about to end, we might, at this point, take up again the discussion of the logical aspects of indeterminacy. However, let us first take advantage of the opportunity that has grown out of the discussion concerning the precision of a measurement in the quantum theoretic field, in order to examine the question in relation to the considerations we put forward about the subject in general (in Sections 7 and 8). We have just pointed out the similarity between relations of indeterminacy and bounded precision in terms of area; more fundamental, however, is the analogy between the case of 'unbounded precision' (considered in Section 7) and the situation presented (following von Neumann) for the measurements of nonquantized quantities (with continuous spectra). These quantities (vN, p. 222) '… could be observed only with arbitrarily good (but never absolute) precision', in contrast to what happens with the '… introduction of an eigenfunction which is "improper", that is, which does not belong to Hilbert space', a procedure which '… gives a less good approach to reality than our treatment here. For such a method pretends the existence of such states in which quantities with continuous spectra take on certain values exactly, although this never occurs.' These critical comments are directed towards procedures that make use of the Dirac function (and they are repeated very frequently). In this connection, I think it appropriate to indicate its relation to the point of view that we are following, both to avoid misunderstanding and to make things clearer. I sympathize with von Neumann's attitude, not when he seems to be inspired by scruples of mathematical rigour and attacking imprecise definitions (because, generally speaking, formal imperfections can always be removed), but when he shows care in not attributing absolute certainty and precision to a quantity without really good reason. I approve even more strongly of the observation (as he adds in note 126, p. 222) that not even attributing X to one of the intervals of the subdivision can be considered as certain, except as an idealization: 'Nevertheless,' he concludes, in an admirably undoctrinaire manner, 'our method of description appears to be the most convenient one mathematically at least for the present.' 'I sympathize', also, in the sense that I would like quantum physics to make room for this elegant example of contraposition; of quantities that are or are not quantized, which are, respectively, precisely measurable or not. In any case, it cannot, of course, be a matter of taste, whether mathematical or philosophical, and if the opposite formulation should, on a closer examination, turn out to correspond more closely to a meaningful physical interpretation then it must be welcomed with open arms.

12 The Relationship with 'Three-Valued Logic'

Let us now go back to three-valued logic and to the related conceptual questions that are raised by quantum mechanics.

So far as the nature of 'three-valued logic' is concerned, we came to the conclusion that the 'three values' correspond well to the requirements of applications – quantum-theoretic or otherwise – but that they do not give rise to a 'logical calculus', because the most meaningful considerations are not connected with operations that could be performed on such 'values'. If one examines the actual situations directly, the preconceived choice of a machinery consisting of formal operations similar to those of ordinary logic does not appear to be appropriate as the unique way of constructing a formulation which is to replace it.

The best proof is provided by the many-valued logic – 'similar to the calculus of probability' – which Bodiou mentions and develops. In order to have the satisfaction of finding a true relationship (in the calculus of probability), he has to put together two things which are falsely defined (in the calculus of probability) and which, on the other hand, cannot be modified if one wants them to be expressed as 'logical functions' (and the final consolation lies in the observation that they only work if one has either probability 0 or 1, in which case formal, two-valued, logic, without probability, is essentially sufficient).[43]

The most important point to be examined is the reason behind the different attitudes (already mentioned in Section 10) of those who regard 'Indeterminism' as a concept specifically and exclusively belonging to quantum physics, and those who see no distinction *of a logical nature* between this case and the world of everyday affairs (although nobody denies the very important and significant differences which derive from the physical and mathematical structures peculiar to the quantum-theoretic set-up).

Von Neumann, in speaking about the 'sort of logical calculus' to which the projection-operators gives rise, says of this calculus that '... in contrast to the concepts of ordinary logic, this system is extended by the concept of "*simultaneous decidability*" *which is characteristic for quantum mechanics*' (vN, p. 253).

It seems that such sentences have no practical implication and, therefore, no actual content. So far as von Neumann is concerned, it may be that he never examined the possibility of examples of a different kind. This is not the case, however, with Reichenbach (as we shall see); indeed, one might think that he was constantly preoccupied with such

43 Perhaps this 'many-valued logic' might be useful in other cases, and in other senses, without reference to probability. For example, by giving a proposition a a certain value $V(a) = k$, if, in a system with a given set of ordered axioms $A1, A2, \ldots, An$, the proposition is decidable (true or false) on the basis of $A1, A2, \ldots, Ak$, but not using only $A1, A2, \ldots, A(k-1)$. The scheme as it stands does not even work in this case, but, in a certain sense, we get closer.

So far as non-modularity is concerned, one can observe that modularity no longer holds in our scheme, $(E|H)$ (or in similar ones), when the 'truth-value' (Void or Indeterminate) is considered greater than 0 (False) and less than 1 (True), and a scale of intermediate values (in various possible senses, e.g. probability), which are considered not comparable with Ø, are inserted between these two values. The most natural convention would be that of taking $\mathbf{P}(E|H)$ as the value, putting $\mathbf{P}(E|H) = \emptyset$ if $H = \emptyset$; possibly using 0^* and 1^* to distinguish the certainly False and certainly True cases (obtaining the partially ordered set of values $0^* < 0 \leqslant p \leqslant 1 < 1^*$, $0^* < \emptyset < 1^*$, Ø not comparable with values $0 \leqslant p \leqslant 1$). It is not clear to me whether this has any connection with the appearance of non-modularity – in many different ways, some not immediate – in the treatments given by von Neumann and Birkhoff, and Bodiou.

dilemmas (like waves versus particles, values before and after etc.). These dilemmas were, instead, clearly resolved by von Neumann, as is shown, for example, by the following remark (note 148, p. 282):

> 'In contrast with this, however, it is to be noted that quantum mechanics derives both "natures" from a single unified theory of the elementary phenomena. The paradox of the earlier quantum theory lay in the circumstance that one had to draw alternately on two contradictory theories (electromagnetic theory of radiation of Maxwell–Hertz, light quantum theory of Einstein) for the explanation of the experience.'

The attitude of Bodiou – see the previous quotation (B, p. 7) – also seems to stem from having overcome these distinctions. On the other hand, it seems an inevitable progression to attain more and more comprehensive views, which remove from their isolation those things which, when they first appeared, seemed abnormal.

What is the difference then, from a logical point of view, between the complementarity or noncomplementarity of measurements in the case of a physicist and in that of the tailor (whom we met above in our trivial example)? Or among the examples given previously – like that of the coin whose next toss could be made by either Peter or John – and an example of a quantum-theoretic nature? It is precisely in the context of such an example that Reichenbach has developed his arguments (R, pp. 145–146, and p. 168), basing himself upon an absolutely rigid division between the indeterminacy of the quantum world and the determinacy of the macroscopic world. This division is so complete that Reichenbach says the following concerning the outcome of *that toss which John might have made* (but which instead was made by Peter). Since it is a question of 'a macroscopic affair, we have in principle other means of testing', by making precise measurements of the state of John's muscles before or after the toss made by Peter, and in many other ways:

> '… or let us better say, since we cannot do it, Laplace's superman could. For us the truth value of John's statement will always remain unknown; but it is not *indeterminate*, since it is possible in principle to determine it, and only lack of technical abilities prevents us from so doing.'

In discussing the merits of the question, one might object that the 'determinism' of the macrocosm – to which Reichenbach makes explicit reference – has a merely static character and this renders completely unpredictable those facts for which numerous microscopic circumstances might prove decisive (not to mention the fact that even the result of a single collision between particles, as recorded on a photographic plate, can cause macroscopic phenomena like the publishing of papers, the holding of lectures and conferences, and endless indirect consequences and repercussions). Moreover, not even Laplace (so far as I know) ever suggested that his 'superman' was capable of predicting not only everything that is going to happen but also what would happen if … something that is not going to happen were to happen. How could it come about that the state of the muscles, and so on, could inform us about the result of the toss that has not been performed (and why not the text of a conversation that has not taken place; the adventures of a journey not undertaken; etc.), rather than informing us directly that it is predetermined that the toss, or the conversation, or the journey, will not take place (or did not take place)? To my knowledge, no-one, even in theological discussions, has ever claimed to have decided whether divine omniscience includes the knowledge of what exactly would have happened to the world conditional on every imaginable hypothesis about the form of Cleopatra's nose (or any other fact, either substantial or irrelevant, concerning the world's history).

In my opinion, however, there is no point in entering into the merits of such questions, physical or metaphysical as the case may be, because logic can only be *neutral* and *anterior* with respect to any contingent circumstance of scientific knowledge, or hypothesis, concerning the world of phenomena.[44] Logic has to be applied to the wider field of everything that is imaginable, and the inevitable circumstance that fantasy is of so little use in extending the field beyond what has already been observed or realized is already too restrictive. Science fiction itself has rarely anticipated reality by more than a few decades. To make use of new ideas or discoveries can be legitimate for the purpose of bringing up to date points of view in logic by including in its domain new areas of what is conceivable, areas which had previously been ignored (and this is what we are attempting to do). The approach, which consisted, instead, in making every logical theory restrictive and ephemeral, by reducing it, moment by moment, to a reflection of current scientific views, would have got things upside down.

Before turning to another topic, it would perhaps be appropriate to clarify certain views on the theme of determinism, given the connection with discussions pertaining to the present theory, and given that we have commented upon it (even though in order to decide that it was not relevant). In my opinion, the attachment to determinism as an *exigency of thought* is now incomprehensible. Both classical statistical mechanics (or Mendelian hereditary) and quantum physics provide explanations – in the form of coherent theories, accepted by many people – of apparently deterministic phenomena. The mere existence of such explanations should be sufficient to give the lie for evermore to the dogmatism of this point of view. What I mean is that the fact that such theories exist, or are conceivable, should be sufficient (no matter if they are wrong, or even if they are merely successful mental constructs, clearly science-fictional in character).

It is a rather different matter to pose oneself similar questions from what one might call a psychological–aesthetic angle, rather than a dogmatic one. As a result of our own individual tastes and habits, each one of us will have a propensity to find one or other of a deterministic or indeterministic formulation of a law or theory more or less simple and convincing. In particular, we evaluate with greater or lesser sympathy (*a priori* – i.e. before some possibly deeper knowledge or examination of the detailed reasons for and against) the ideas which tend to characterize probabilistic-type quantum theory as merely a partial explanation, unsatisfactory and provisional, and requiring replacing sooner or later by something deterministic.

Personally, I am of the opinion that nothing should ever be excluded *a priori*: tomorrow's notions will almost certainly be as inconceivable for us today as today's notions would have been for a man of the nineteenth century, or for Neanderthal man. This is, however, a distant prospect; the foundations of physics are those we have today (perhaps for many decades, perhaps centuries) and I think it unlikely that they can be interpreted

44 On the other hand, this has been perfectly expressed elsewhere by Reichenbach himself.

(or adapted) in deterministic versions, like those that are apparently yearned for by people who invoke the possible existence of 'hidden parameters', or similar devices. I hold this view not only because von Neumann's arguments against such an idea seem to me convincing (vN, pp. 313–328), but also because I can see no reason to yearn for such a thing, or to value it – apart from an anachronistic and nostalgic prejudice in favour of the scientific fashion of the nineteenth century. If anything, I find it, on the contrary, distasteful; it leaves me somewhat bewildered to have to admit that the evolution of the system (i.e. of its functions ψ) is deterministic in character (instead of, for example, being a random process) so that indeterminism merely creeps in because of the observation, rather than completely dominating the scene. This can actually lead one to search for some meaning which makes the function ψ objective, although this notion is the very least suited to appear to be capable of such a transformation.

In any case, for what concerns us as human beings, interested in foreseeing the future with some degree of confidence on the basis of our scanty, imprecise and uncertain knowledge of the present and the past, all arguments about determinism are purely academic and have no more meaning than would a discussion about the number of angels that can dance on the head of a pin. No matter how the world's history develops, nobody could disprove either the assertion that everything is determined by the past through iron laws (but we can foresee either nothing or very little because we are too ignorant both of the past and of the laws), or the assertion that everything occurs 'by chance' (and this does not exclude the possibility that 'by chance' things might develop according to some 'law'). In the final analysis, it seems to be of very little consequence or assistance to us whether we take up a position for or against the plausibility of the hypothesis that Laplace's superman could work out the entire future *if only he knew the entire present in every detail.* Such a statement, in the sense we have just examined, must, in fact, be said to be neither true nor false, but instead indeterminate, the hypothesis being, without any doubt, illusory, and therefore false.

13 Verifiability and Distorting Factors

As the final part of our survey of the various factors that are important when we attempt to determine and verify the outcome of an event, it remains to consider the most troublesome of them. These are the factors which, for reasons relating to the individuals involved, or to their self-interest, are capable of modifying the outcome, or of influencing its verifiability, or of simply raising doubts about the possibility of such distortions.

Many examples of this are well-known, and we shall just quote them without having anything useful to say about overcoming the difficulties. The deepest discussion (which may well be new) will centre, however, on the events of the three-valued logic illustrated in Sections 10–12 above, where it seems impossible to give a complete definition without encountering similar difficulties regarding possible distortions. Let us begin, however, with the most well-known and obvious cases.

These are events for which the will of the individual concerned enters in directly (and, in this respect, there is nothing to distinguish this case from that of events which depend on animals, or on other natural factors). There is a difference – a distorting factor – when such a will can be influenced by facts which are objects of our study, and which therefore alter this very object of study. This happens in evaluating a probability if the circumstances upon which the evaluation itself is based are modified by the evaluation, or the knowledge of this evaluation, or a contract drawn up on the basis of this evaluation, and so on.

This much is true: although we are again speaking of probability before the appropriate time, we have to do so in order to present the examples; the object under consideration is, however, the difficulty of avoiding the problems by means of detailed specifications in the description of an event.

The evaluation of the probability of an event can influence its occurrence. If someone, at some given moment, perhaps because of a vague feeling, or even for no reason at all, considers the danger of a traffic accident to be higher than usual, he will try to be more careful, and the risk will diminish. If, on the other hand, we are dealing with an event whose positive outcome is desired – like succeeding in a business deal, an examination, or a race – it can happen that a greater feeling of confidence leads to one being in a better position to succeed.

The knowledge of someone else's evaluation of probability can have a marked snowball effect, as a result of the confidence that tends to be placed in the opinions of experts. If in circles which are considered well-informed the expectations are pessimistic (or optimistic) and an increasing number of people, when informed of these opinions, behave as if they correspond to reality, the expectations will end up by being borne out by reality – even if they were initially without foundation.

Nevertheless, the most direct example is provided by the influence of a person's self-interest on the outcome of events. In the case of insurance, it can lead to faked or fraudulent accidents; but we are still dealing here with the kind of case which is, to some extent, identifiable. Much worse (from a logical point of view, since it constantly eludes one's grasp) is the effect of the insufficient precautions that an individual might take, knowing that he is insured. Similar influences are at work if a prize is attached to the occurrence of an event (e.g. an additional bonus for each goal, either for the player who scores, or for the whole team to share out), or even if it arouses admiration, or merits reproach.

It is even easier for a person to have an influence on the process of verification rather than on the outcome itself. An individual who is interested in proving that an event has occurred will devote a lot of energy to obtaining information to this end, and will take care to collect the necessary documentation and to send it to the appropriate authority. On the other hand, someone interested in concealing such news will be more or less negligent, or may even attempt to suppress it, or to destroy the evidence.

In order to avoid all this, one should provide a description of the event which is sufficiently detailed to preclude the possibility of distortion. In fact, the clauses of an insurance policy abound in detailed specifications of the obligations of the person concerned, the risks excluded, and so on (although it is clearly not possible to extend the specification beyond those cases which are easiest to define and to pick out[45]).

An even more entangled situation is to be found in the theory of games. In the simplest case, one has two players, each of whom must make a decision (without knowing the decision of his opponent), and the result (one player's gain, the other's loss) depends on the two decisions made. Each would like to know the decision of the other in order to

45 A discussion, together with useful examples, can be found in H.M. Sarason, 'Come impostare e applicare le statistiche assicurative', *Giorn. Ist. Ital. Attuari*, I (1965), pp. 1–25.

adjust his own decision accordingly; not knowing it, he could evaluate the probabilities of the various decisions the other might make, and in order to do this he would need to go through a similar reasoning process by putting himself in the other's shoes.

This and other more complicated situations are objects of study in games theory. But all the aspects of distorting factors that we have mentioned so far are only intended as remarks in passing, merely to put the reader on guard against the difficulties one encounters when dealing with cases where they arise (the difficulties may or may not be serious, but they are virtually impossible to eliminate).

Above all, these examples serve as an introduction, in order that it should not appear (misleadingly) to be a rather special feature which arises when one delves more deeply into the study of 'three-valued' events. There is, in fact, a particular, novel feature in this case, but it arises later, and is not related to the distorting factor (which derives from the choices that can affect verifiability). We shall examine this latter aspect first.

A conditional event $E|H$ presents no problem of this kind if E and H turn out to be known with certainty – as true or false – within the time and manner specified. In fact, if we think in terms of having made a bet – our guideline – we will then know, without any room for doubt, that it is called off if H turns out to be false, won if both H and E turn out to be true, lost if H but not E turns out to be true. But what if H or E, or both, turn out to be nonverifiable (in some preassigned manner; for example, within a given time period during which the bet has to be decided)? As a first step, we must decide what is to happen to the hypothetical bet in such circumstances. It seems natural – and, in any case, this is what we shall do – to make the following convention: it is either won or lost, respectively, only in the cases of H and E true, H but not E true, respectively; it is called off both if H is false, and if H is indeterminate, and also if H is true but E is indeterminate. In formal terms, considering E and H as three-valued events, $E = E'|E''$, $H=H'|H''$, the conditional event $E|H = (E'|E'')|(H'|H'')$ would correspond (in Reichenbach's terminology) to *quasi-implication* (as introduced by him), with the following truth table (in Reichenbach's notation, $E|H$ corresponds

$E\|H$		H: T	I	F
E	T	T	I	I
	I	I	I	I
	F	F	I	I

to $H \ni E$). In terms of the four simple (two-valued) events E', E'', H', H'', this becomes

$$E\,|\,H = (E'\,|\,E'')\,|\,(H'\,|\,H'') = \begin{cases} 1 & (T) \text{ if } E'E''H'H'' \\ \varnothing & (I) \text{ if } \sim(E''H'H'') \\ 0 & (F) \text{ if } \tilde{E}'E''H'H''; \end{cases}$$

in other words,

$$E|H = (E' \mid E'')|(H' \mid H'') = E' \mid (E''H'H'').$$

Distorting factors enter in here, too, as soon as one allows the possibility that somebody might influence the outcome, or the knowledge of the outcome, of $E|H$. The particular case of greatest specific interest is that in which H represents the performing of the experiment – or, more usually, experiments – from which information about the outcome of E is drawn (either in fact, or potentially). This covers the cases of all measurements and experiments in both classical and quantum physics, and of all the investigations that are appropriate for the ascertaining of the truth of any assertion concerning practical matters. This situation arises most clearly when E consists just of the result of an experiment which must be expressly performed (e.g. H = the toss of a coin, or the launch of a satellite, and E = Heads, or entering into orbit, respectively). In such a case, it would make no sense at all to enquire whether E was true or false without assuming that H were true; but the case in which E is thought of as being true or false independently of an experiment H for ascertaining it no longer appears different when one is concerned with the actual ascertainment of E. We can very well imagine that E = A.N. Other has gone down with a certain illness, or E = the residue of a substance contains poison, are statements that are true or false in themselves, independently of the fact of our knowing whether they are true or false. If we, or anyone, wish to know whether E *is* true or *is* false (and not merely to say that it is one or the other) then E should be replaced by $E|H$, where H denotes the performing of an act leading to its ascertainment. We could say, for example, that H = A.N. Other undergoes tests to establish whether or not he is infected with the given disease, or H = the residue of the substance analysed in order to ascertain whether or not it contains poison, and then E = the outcome is positive. But what tests and analyses should be performed?

Let us exclude the possibility that an experiment (e.g. the tests or analyses mentioned in the above examples) could give a wrong answer: this would by no means be absurd, because experiments concerning facts related to the one we wish to ascertain can only lead us to increase or decrease the probability we attribute to it. Everything stems from our convention of considering that a question has been answered only if it is certain, and we call E indeterminate if the ascertainments that have been made have not proved sufficient to resolve the doubt. (Just as, in the case of 'insufficient evidence', it would be inadmissible to claim that a suspect is both guilty and not guilty.)

However, it is only rarely that by performing an experiment H one obtains an answer with certainty. What usually happens (at least in cases which are sufficiently complicated for it to be worthwhile to apply considerations of this kind) is that H may give an answer (in which case it is an exact answer), but, on the other hand, may not (and then E remains indeterminate). To be precise, H should not simply denote the performance of some given experiment, but rather the successful performance of it (in the sense that, with respect to E, the answer is either YES or NO, and not MAYBE). If we wished to split hairs, we could put $H = K'|K$, where K denotes the experiment in the sense of its performance, K' the fact that K was a success, and thus

H is the successful performance of K (i.e. the hypothesis that ensures the ascertainment of the truth or falsity of E).[46]

In general, however, there will be no one unique experiment K which we can (or cannot) perform in order to ascertain E. There will exist various possibilities $K_1, K_2, K_3,\ldots$ (and even if there were only one 'type' of experiment available, we could always vary the time, the apparatus, or the experimenter etc.) and they might or might not be compatible for various reasons (and possibly not repeatable in the case of failure), ranging from physical incompatibility to contingent limitations (e.g. lack of time, apparatus or available personnel, funds, raw materials etc.). In order not to further complicate the notation, we can suppose that the list given by the K_i includes not only the individual experiments (e.g. let $K_1, K_2,\ldots, K_{38}$ denote the performance of just one of 38 possible different experiments), but also all possible combinations or strategies (e.g. the one consisting in first performing K_5, K_{19} and K_{22}, and then, if none of these succeeds, K_7, and, if this is still not sufficient, K_9 and K_{31} together, and then stopping whatever happens, is a strategy which will be denoted by a number greater than 38 – e.g. by K_{728}). For each K_i, K_i' will mean that K_i was successful; that is that K_i succeeded in establishing whether E was true or false. In the case of an individual experiment, K_2, say, K_2' will mean that this experiment was successful; in the general case, for example for K_{728} ,K_{728}' will mean that at least one of the component experiments of the strategy was successful (and – in the case of a sequence of experiments, as in the example of K_{728} – the experiments following will then not even be performed). By going into more and more detail (like the time and manner of performing the experiments, possible repetitions etc.), the number of distinct strategies could be increased without limit. So far as our notation is concerned, however, it is simply a question of extending the list of the K_i.

In this way, the problem of ascertaining E is translated, in practice, into the problem of ascertaining one of the conditional events $E|K_i$, depending on the choice of K_i, which is arbitrary within the limitations imposed (of time, money etc.). And it is thus that the arbitrariness has its effect on the verification of E. In extreme cases, there might be one-way experiments; that is experiments which either prove that E is true, or prove nothing (or vice versa). Suppose that one experiment shows whether or not a given liquid is pure water, and another experiment shows whether or not it contains strychnine. If the question is whether or not the liquid is poisonous, the first experiment can only return a negative answer, and the second only a positive one (because to know that it is not pure water, or that it does not contain strychnine, neither proves nor disproves the presence of poison). Even without taking these extreme cases into consideration, any method might present, by its very nature (and taking previous experiences into account, according to the evaluation of each individual), different characteristics in its functioning, and different probabilities of breaking down, depending on whether E is true or false.

Up to this point, we have merely been dealing with cases involving distorting factors, just like the others considered previously (even if these cases deserve special attention because they are less obvious than most other examples). The most important specific

46 Only thus can one avoid having to consider E itself (as well as $E|H$, which becomes, in this notation, $E|K$) as a three-valued event (indeterminate, notwithstanding the – unsuccessful – performance of the experiment). We have our doubts about the actual utility of such notation, other than for a once and for all explanation, and we avoid insisting on, or taking up a position, regarding the desirability of more or less logically perfect forms of notation.

factor in the present case is, however, quite a different one, which – as we mentioned at the beginning – only 'arises later', after the analysis given above of the experiments K_i, their successes K_i', and the consequent realization of the corresponding hypothesis H_i.

The new factor is the following: to be realistic, one should also substitute E_i for E. Let us explain right away what we mean by this. If we perform the experiment K_i, its success K_i' does not give us directly the answer 'E is true', or 'E is false'; it does not make directly visible to us the fact that we wish to affirm or deny by these phrases. Neither, if we are dealing with the more general question of measuring a quantity, does it enable us to realize what the value is by making it visible or tangible. The answer reduces to a signal (a movement, a light, a noise, a colour etc.; in the case of quantities, the position of a pointer on a dial, the reading of a counter, the height of a column of mercury etc.). For an event, we shall have one of two signals, E_i or $\tilde{E}_i$, as possible outcomes of the experiment K_i (as well as the absence of any answer at all – or, if one prefers, $\emptyset$, or $\tilde{K}_i'$), and these may differ from experiment to experiment. *But*, it could be argued, *this is irrelevant, because we know that they correspond to E being true, or E being false.*[47]

Agreed ..., but what does this mean exactly? The last sentence, so simple, clear and straightforward, is admirably suited to an hypothesis which is equally simple and clear; the hypothesis which assumes that one of the many experiments has been taken as the *definition* of E (i.e., if the one chosen is K_{13}, then E means E_{13}, or, better, $E_{13}|K_{13}$), and that this experiment is always possible and always successful. It follows that the statement that E is true because a different experiment K_4 has given the signal E_4, can be strengthened by the remark 'since it is certain that the answer E_{13} – that is E – will be obtained if we perform experiment K_{13}, it is unnecessary to do so, because of the trouble, expense and so on; but, if you do not believe it, try it, and you will see!'. In this way, for example, if I derive the height of a distant tower by trigonometric methods, or by observing how long it takes for stones dropped from the top to reach the ground, I can say to someone who does not believe it 'go up and measure it'. And one might allow that the argument is considered in general to be valid, even when the invitation becomes rather less realistic (distance from the centre of the earth, distance between two stars, or two galaxies). But what if someone does not believe it?

In the previous hypothetical case, there was a criterion which could appropriately be assumed as a definition, both because of its meaning and because it could always be applied (at least in principle) with a guarantee of success. In particular, it could be applied to statements about things that are not directly observable and possess disconcerting properties (as in the case of 'waves' and 'particles'). What do we do in the absence of such a criterion? We could define all the experiments K_i (and the simple ones will suffice, it is not necessary to consider strategies) by means of the respective answering signals and observe that, in any case, no matter which K_i are applied, and no matter how many of them, in all successful cases they give a concordant answer; either always E_i, or always $\tilde{E}_i$. This, practically speaking, assures the meaning and uniqueness of the notion related to E (or, more generally, to a quantity), provided that the coincidence of answers for any two methods, K_i and K_j, could be verified experimentally by applying both of them in precisely that same situation (or at least indirectly, by means of a chain of equivalences, each link of which could be verified experimentally).

47 See the discussion of H. Jeffreys that we have already quoted (Chapter 11, the end of 11.1.1).

But what if we come across cases where it is not possible to perform more than one experiment in a given situation? Any statement of the form 'having observed the outcome E_i of the experiment K_i, we know that *had we performed* the experiment K_j we *would have* obtained E_j', is entirely without content, since the assumption is false. It is the same situation as the one we already considered, albeit light-heartedly, at the beginning of Chapter 4 (Section 4.1), when we asked 'whether or not it is true that had I lived in the Napoleonic era and had participated in the Battle of Austerlitz I would have been wounded in the arm'.

It might help to place one's confidence in a more speculative kind of generalization, such as one makes in passing from the direct verification of indirect measurements of length on the 'human' scale to admitting the same thing for inaccessible distances. The generalization required would have to admit the coherence and validity (justified by numerous indirect proofs) of the entire set of concepts, arguments and calculations which constitute the scientific view of the world.

It is a fact, however, that, so far as the 'ordinary man in the street' is concerned, the only reason he believes in these things is a lack of appreciation of the fact that they are more abstruse and delicate than he imagines. The logical situation for him is, under the worst assumptions, the following: he is given the explanation that the fact of E being true (to fix ideas, think of E as an event on which he would like to bet) can be verified in one and only one of the many possible ways provided by performing an experiment K_i and receiving a corresponding answering signal E_i (the choice being made by the experimenter; in any case, *this tells him about the same fact*). But this leading statement conveys nothing to the man in the street, who has no idea of what 'fact' it is that one is dealing with. (For the scientist, too, it is very much an intellectual conviction; but we are not interested in him.) The man in the street only knows, naturally enough, that he might bet on the outcome of an experiment whose choice is in the hands of his opponent. He may therefore think (in theory, of course – not in practice, because it is not nice to be suspicious) that the choice will be made to his disadvantage: for example he may end up trying to draw a white ball from an urn containing only black balls, this being one of the possible choices open to his opponent.

Leaving these more or less picturesque illustrations aside, it would seem that the conclusion – a negative and disturbing one – cannot be other than the following: *one does not succeed in giving an operational meaning to a statement E (or to a quantity X) by means of a collection of statements $E_i|H_i$, which do have operational meaning, without introducing the statement that all the obtainable results E_i are necessarily conformable (and this does not have an operational meaning if the H_i are incompatible)*.

14 From 'Possibility' to 'Probability'

The logic of certainty only distinguishes events which are either true or false, and which can only be *possible* (uncertain) rather than certain or impossible[48] (for us, in our more or less temporary state of ignorance). We have discussed questions of a critical nature

48 We do not have to worry about 'indeterminacy', considering it (see Sections 9–12) reducible to the case of two-valued logic by means of conditional events.

by remaining within this ambit as a preliminary exploration of the field into which we have to introduce and apply the logic of the probable.

The time has now come to deal with the main critical questions, which specifically concern the subject of direct interest to us: the theory of probability.

We do not want to repeat ourselves by going back to the beginning and starting from scratch: we have already dealt with many questions in the text, and many comments were necessarily made as we developed our approach. It will be more appropriate to refer back to these, to draw the threads together, and to go more deeply into them, in order to provide a synthesis and, finally, what will, hopefully, turn out to be a sufficiently integrated view of the entire subject.

Let us begin by sketching a broad outline, including both those topics for which we shall limit ourselves to recalling our previous remarks – or just adding the odd word here and there – and those which we shall take up again later because they require further analysis or more thorough discussion.

Without further comment, we shall take as axioms those already established as the basis of our subjectivistic formulation. This will in no way prejudice the (technically neutral) possibility of comparison; our starting point, in fact, makes comparison easier, because it represents the minimal set of conditions common to all formulations. The subjectivistic formulation, as we have said repeatedly, is, in fact (and deliberately so), the *weakest* one; its only requirement is coherence, and in no way does it seek to interfere with an individual's freedom to make an evaluation by entering into the merits of it on some other grounds.

In discussing these concepts, we shall provide a comparison with other points of view, which differ in various respects (in the interpretation of the notion of probability, the mathematical details and the qualitative formulation).

Those interpretations of the notion of probability in a (would-be) objective sense that are based on symmetry (the classical conception; equally likely cases), or on frequency (the statistical conception; repeated trials of a phenomenon), provide criteria which are also accepted and applied by subjectivists (as, to a considerable extent, in this book). It is not a question of rejecting them, or of doing without them; the difference lies in showing explicitly how they always need to be integrated into a subjective judgment and how they turn out to be (more or less directly) applicable in particular situations. If one, instead, attempts to force this one or that one into the definitions, or into the axioms, one obtains a distorted, one-sided, hybrid structure.

The mathematical details remain those that derive from the positions we adopted concerning zero probability, countable additivity and the interpretation of asymptotic laws (points which we have already encountered, and commented on, many times). In this regard, we shall have to consider many further points, which we glossed over in Chapters 3, 4 and 6, in order not to overcomplicate the exposition (prematurely), and to add some details concerning a number of new features. These considerations, together with some others, will enable us to sort out, and comment upon, the differences between the axiom system we have adopted here, and that given by Kolmogorov (1933), the formulation which, broadly speaking, has been adopted by most treatments of the last few decades.

Finally, under the heading of 'qualitative formulations', we will have to mention two separate topics. The first concerns the possibility of starting from purely qualitative axioms – that is in terms of comparisons between probabilities of events (this one is more probable than that one etc.) – without introducing numerical probabilities, but eventually arriving at them by means of comparisons of this kind. The second deals with the thesis that several authors have recently put forward, namely that probabilities are intrinsically indeterminate. The idea is that instead of a uniquely determined value p one should give bounds (upper and lower values, p' and p''). That an evaluation of probability often appears to us more or less vague cannot be denied; it seems even more imprecise, however (as well as being devoid of any real meaning), to specify the limits of this uncertainty.

15 The First and Second Axioms

The entire treatment that we have given was based on a small number of properties, which were justified in the appropriate place in the text as conditions of coherence. In order to develop the theory in an abstract manner, it will now suffice to assume these same properties as axioms.

There will be two axioms (the first and the second) dealing with previsions and a third dealing with conditional previsions. The third one – which is needed in order to extend the validity of the first two to a special case – will be dealt with later (Section 16); we concentrate for the time being on the first two.

Axiom 1 *Non-negativity*: if we *certainly* have $X \geqslant 0$, we must have $\mathbf{P}(X) \geqslant 0$.

Axiom 2 *Additivity* (finite):

$$\mathbf{P}(X+Y)=\mathbf{P}(X)+\mathbf{P}(Y).$$

From these it also follows that

$$\mathbf{P}(aX)=a\mathbf{P}(X), \quad \inf X \leqslant \mathbf{P}(X) \leqslant \sup X,$$

as well as the (Convexity) condition, which includes Axioms 1 and 2:

(C) *any linear equation* (*or inequality*) *between random quantities* X_i *must be satisfied by the respective previsions* $\mathbf{P}(X_i)$; in other words,

$$\text{if we } \textit{certainly} \text{ have } c_1X_1+c_2X_2+\ldots+c_nX_n=c \left(\text{or} \geqslant c\right)$$

$$\text{then } \textit{necessarily } c_1\mathbf{P}(X_1)+c_2\mathbf{P}(X_2)+\ldots+c_n\mathbf{P}(X_n)=c \left(\text{or} \geqslant c\right).$$

By taking differences, (C) can be written in an alternative form:

(C') *No linear combination of* (*fair!*) *random quantities can be uniformly positive*; in other words, the $\mathbf{P}(X_h)$ must be chosen in such a way that whatever be the given $c_1, c_2,\ldots, c_n$, *there does not exist a* $c > 0$ *such that*

$$c_1\left(X_1-\mathbf{P}(X_1)\right)+c_2\left(X_2-\mathbf{P}(X_2)\right)+\ldots+c_n\left(X_n-\mathbf{P}(X_n)\right)\geq c$$

certainly holds.

We could put forward as a further (possible) axiom one which consists in excluding the addition of other axioms; that is one which considers *admissible*, as prevision-functions $\mathbf{P}$, all those satisfying Axioms 1 and 2, or equivalently, condition (C).[49] On the other hand, this is implicit, since nothing is said to the contrary. In any case, we shall say that every function $\mathbf{P}$ satisfying Axioms 1 and 2 is *coherent.*

As we have already mentioned (Chapter 3, 3.10.7), a coherent function $\mathbf{P}$, defined on some given set of random quantities X (an arbitrary set, in general infinite), can always be extended, preserving coherence, to any other random quantity, X_0, say. From any inequality of the form (C'), one can obtain, by solving it with respect to one of the summands (let us assume $c_0 = \pm 1$ and take it to be the one corresponding to X_0; were this not the case, it suffices to divide through by $|c_0|$), an inequality for $\mathbf{P}(X_0)$ of the form

$$\mathbf{P}(X_0) \leqslant \inf\left\{X_0 + \sum_{h=1}^{n} c_h\left(X_h - \mathbf{P}(X_h)\right) - c\right\} \qquad \left(\text{or} \geqslant \sup\{\ldots\}\right).$$

As a result, we obtain $x' \leqslant \mathbf{P}(X_0) \leqslant x''$, where x' denotes the greatest lower bound and x'' the least upper bound. If $x' = x''$, the extension will turn out to be uniquely defined;

$$\mathbf{P}(X_0) = x' = x'',$$

that is $\mathbf{P}(X_0)$ will be determined by the values given over X. If $x' < x''$, the admissible values for $\mathbf{P}(X_0)$ will consist of all those in a closed interval (as is obvious by convexity). The extension would be impossible if $x' > x''$, but this is ruled out by the observation that there would then exist a linear combination $X_0 + \sum_i c_i\left(X_i - \mathbf{P}(X_i)\right)$ always $>x'$, and another one

$$X_0 + \sum_j c_j\left(X_j - \mathbf{P}(X_j)\right)$$

always $<x''$; their difference ($\Sigma_i - \Sigma_j$; X_0 cancels out) would then turn out to be $> x' - x'' > 0$. But this would mean that there was a contradiction of (C') already contained in X, contrary to the hypothesis.

It follows immediately from this that one can always define a $\mathbf{P}(X)$ for all the X belonging to an arbitrary set of random quantities (in particular, one can always define a $\mathbf{P}(E)$ for every event in an arbitrary collection of events – for example those corresponding to all subsets of a given space), even assuming $\mathbf{P}(X)$ as already assigned in some given field, and extending it. It is sufficient, as we have done here, to carry out the extension for new X one at a time, by means of *transfinite induction* (assuming, of course, the Zermelo Postulate, in order to well-order the X_h; the indices, h etc., will be transfinite ordinals). One has to be a little careful that nothing goes wrong for the X_k which have no

49 Note that we are not dealing here with the basic issue of whether under given circumstances all the coherent evaluations $\mathbf{P}$ are admissible (subjectivistic conception), or whether only one of them corresponds to reality (objectivistic conceptions). For the objectivist also, it is a question of knowing which $\mathbf{P}$ are formally admissible (e.g. the $\mathbf{P}$ which he can adopt when he has the information he is now lacking – about composition of urns, frequency of statistical phenomena etc.), or even that he judges to be possible with respect to the abstract scheme without knowing which concrete events are represented by the symbols E_1, E_2 etc. On the other hand, this is the attitude adopted by the supporters of all points of view when they are faced with the notion of '(abstract) probability space'.

antecedent (such as X_ω, where ω denotes, as usual, the first ordinal which comes after the natural numbers).[50] In our case, however, the contradiction would derive from the comparison between two *finite* linear combinations and should have occurred at the last of the steps corresponding to the X_h which appear (and the fact that there are an infinite number of steps between this X_h and our X_k does not enter into the argument).

Let us return now to the problem of the extension, in order to consider when it turns out to be uniquely defined. One obvious case is that of a random quantity X_0 linearly dependent on those of the original field X; that is belonging to the linear space L generated by the X belonging to X. In this case, the uniqueness of the extension holds for any **P**.

Condition (C), however, reveals what the situation is in terms of a particular **P**. Instead of linear relations, we have, in general, linear inequalities, $\sum_i c_i X_i \geq c$, which, solved in terms of X_0 (as above for $X_0 - \mathbf{P}(X_0)$), give random quantities X' and X'', linear combinations of random quantities belonging to the field X (and hence belonging to L), which bound X from below and above: $X' \leqslant X$ and $X \leqslant X''$, respectively. We observe that the problem is the same one that we already encountered in a special case (Chapter 3, 3.12.4), and by passing, as here, to the general and abstract case, we also reached essentially the same conclusions. As we vary **P** (defined over X, and hence on L, and, in particular, for X' and X''), the X' for which $\mathbf{P}(X')$ is a maximum, $\mathbf{P}(X') = x'$, will also vary (or, if x' is an upper bound rather than a maximum, the X' to be chosen in order to obtain $\mathbf{P}(X')$ arbitrarily close to x' will vary): similarly for X''. Having chosen X and X'' in this way, we have $X' \leqslant X \leqslant X''$, with $\mathbf{P}(X'' - X') = x'' - x'$ (or $x'' - x' + \varepsilon$, with $\varepsilon > 0$ arbitrary, in the case when they are not the maximum and minimum). In general, therefore, one has a uniquely defined extension if upper and lower bounds of X exist for which the difference $\Delta = X'' - X' \geqslant 0$ has prevision $\mathbf{P}(\Delta) = 0$, or such that $\mathbf{P}(\Delta) < \varepsilon$ (for arbitrary, fixed $\varepsilon > 0$).

In order to denote what can be said about the probability (or prevision) outside some given linear space L in terms of the prevision function **P** defined over it by the evaluations of probability (or prevision), it is convenient to use the same notation (*mutatis mutandis*) as we used in Chapter 6, 6.4.4.

We thus denote by

$$\mathrm{P}_{\mathrm{L}}^{-}(X) = x', \quad \mathrm{P}_{\mathrm{L}}^{+}(X) = x'', \quad \mathrm{P}_{\mathrm{L}}^{\pm}(X) = x,$$

the minimum and maximum (previously indicated in the text by x' and x'') of the values $\mathbf{P}(X)$ which are compatible with the knowledge of **P** over L, and, respectively, their common value (if they are equal).

We shall say more about this – for other reasons – in Section 19.4.

50 Lebesgue measure, too, can be extended, preserving countable additivity, to an arbitrary non-measurable set, and hence to an arbitrary number of such sets, one at a time. In this case, however, an infinite number of steps can lead to a contradiction without any single step doing so (in the same way as a convergent series remains such if we replace the 1st, 2nd, 3rd, …, terms with 1, and so on for any finite number, but not if we replace an infinite number of terms).

16 The Third Axiom

Conditional probabilities $\mathbf{P}(E|H)$, or conditional previsions, $\mathbf{P}(X|H)$, are expressible, in cases where H has nonzero probability, in terms of the unconditional probabilities by means of a formula which, in an abstract, axiomatic treatment, can be taken as a *definition*:

$$\mathbf{P}(E\,|\,H)=\mathbf{P}(EH)/\mathbf{P}(H), \qquad \mathbf{P}(X\,|\,H)=\mathbf{P}(XH)/\mathbf{P}(H).$$

In this case, there is nothing much to add, apart from noting that here, too, an extension (in the sense of $\mathbf{P}_{\mathrm{L}}$) gives rise to an interval of indeterminacy:

$$\mathbf{P}_{\mathrm{L}}^{-}(X\,|\,H)\leqslant\mathbf{P}(X\,|\,H)\leqslant\mathbf{P}_{\mathrm{L}}^{+}(X\,|\,H).$$

To see this, suppose that $\mathbf{P}_1$ and $\mathbf{P}_2$ are two extensions of $\mathbf{P}$ as given over L, and that these give to XH and H the values

$$\mathbf{P}_1(XH)=x_1,\quad \mathbf{P}_2(XH)=x_2,\quad \mathbf{P}_1(H)=h_1,\quad \mathbf{P}_2(H)=h_2.$$

In addition to $\mathbf{P}_1$ and $\mathbf{P}_2$, their convex combinations,

$$\mathbf{P}_\lambda=\lambda\mathbf{P}_1+(1-\lambda)\mathbf{P}_2=\mathbf{P}_2+\lambda(\mathbf{P}_1-\mathbf{P}_2)\ (0\leqslant\lambda\leqslant1),$$

will also be extensions of $\mathbf{P}$, and will give

$$\mathbf{P}_\lambda(X\,|\,H)=\mathbf{P}_\lambda(XH)/\mathbf{P}_\lambda(H)=\frac{x_2+\lambda(x_1-x_2)}{h_2+\lambda(h_1-h_2)}.$$

Since the denominator does not vanish (for $0\leqslant\lambda\leqslant1$; or at most at one of the end-points if one of the h_i is zero, a case that we shall not consider now, however), the hyperbola increases or decreases monotonically between the extreme values

$$\mathbf{P}_1(X\,|\,H)=x_1/h_1 \quad\text{and}\quad \mathbf{P}_2(X\,|\,H)=x_2/h_2.$$

In the extension, the set of possible values for $\mathbf{P}(X|H)$ is thus an interval as asserted.

If $\mathbf{P}(H)=0$, we have a new situation. Does it make sense to consider this case? And, if so, for what purpose? If one were to take the formula, with $\mathbf{P}(H)$ in the denominator, as the actual, unique definition of conditional probability and prevision, then the concept, in this case, would become meaningless. If the meaning were to be assigned in some other, direct, way – for example (as was done in Chapter 4, in line with the subjectivistic point of view), by means of conditional bets – then the meaning would be retained.

But the theorem which expresses coherence, connecting it to the nonconditional $\mathbf{P}$ (the theorem of compound probabilities), no longer holds (and neither does the criterion of coherence) if its formulation (Chapter 4, Section 4.2) has to be in terms of the existence of a 'certainly smaller' loss. In order to extend the notions and rules of the calculus of probability to this new case, it is necessary to strengthen the condition of coherence by saying that *the evaluations conditional on H must turn out to be coherent conditional on H* (i.e. under the hypothesis that H turns out to be true). This is automatic if one evaluates $\mathbf{P}(H) \neq 0$, in which case we reduce to the certainty of a loss in the case of incoherence. The loss for $\tilde{H}$ (Chapter 4, Section 4.3) is, in fact, the sum of the squares of $\mathbf{P}(H)$ and $\mathbf{P}(EH)$; but if $\mathbf{P}(H)$, and therefore $\mathbf{P}(EH)$, are zero, this loss is also zero in the case $\tilde{H}$ (which has probability = 1, and is, in any case, possible).

Although this strengthening of the condition of coherence might seem obvious, we had better be careful with it. There are several other forms of strengthening of conditions, often considered as 'obvious', which have consequences that lead us to regard them as inadmissible. In this case, however, there do not seem to be any drawbacks of this kind; moreover, the 'nature' of the strengthening of the condition seems more firmly based on fundamental arguments (rather than for conventional or formal reasons, or for 'mathematical convenience') than others we have come across, and to which we shall return later. In any case, we propose to accept the given extension of the notion of coherence, and to base upon it the theory of conditional probability, without excluding, or treating as special in any way, the case in which one makes the evaluation $\mathbf{P}(H) = 0$.

If we wish to base ourselves upon a new axiom, we could express it in the following way:

Axiom 3 *The conditions of coherence* (*Axioms 1 and 2*) *must be satisfied, also, by the* $\mathbf{P}_H$ *conditional on a possible H*, where

$$\mathbf{P}_H(E) = \mathbf{P}(E \mid H), \qquad \mathbf{P}_H(E \mid A) = \mathbf{P}(E \mid AH)$$

is to be understood.

This means that $\mathbf{P}_H$ is the prevision function that we may have ready for the case in which H turns out to be true, and the axioms oblige us to make this possible evaluation in such a way that if it is to have any effect it must be coherent. This is implicit in the previous definition if one makes the evaluation $\mathbf{P}(H) \neq 0$. Axiom 3 obliges us to behave in the same way, simply on the grounds that H is possible and we might find ourselves actually having to behave according to the choice of $\mathbf{P}_H$ – even if, in the case in which we attribute probability 0 to the hypothesis H, the sanction provided by the losses does not apply outside of the case $\tilde{H}$.

Axiom 3 permits us to define the ratio of the probabilities of two arbitrary events – even if they have zero probabilities – in the manner already introduced in Chapter 4, 4.18.2. In Section 18.3, we shall expressly return to the topics concerning zero probabilities, topics previously dealt with in Chapter 4, 4.18.3–4.18.4.

17 Connections with Aspects of the Interpretations

The axioms of an abstract theory are, as such, arbitrary and independent of this or that interpretation (at this level, interpretations do not, strictly speaking, exist; or, to put it a little less strongly, one might say that they are ignored).

It goes without saying, however, that the choice of axioms is influenced by the interpretation they will have when the theory is applied in the field for which it has, in fact, been constructed, and on which one would like it to turn out to be adequately modelled.[51]

51 As someone rather neatly put it – Frechet attributes the remark to Destouches – a book which starts off with axioms should be preceded by another volume, explaining how and why these axioms have been chosen, and with what end in view.

In the case of the theory of probability, any judgement about the adequacy of the axioms depends on one's concept of probability and, in addition to the subjectivistic concept, which we have adhered to throughout, we shall also have to consider the 'classical' and 'statistical' concepts.

From the subjectivistic point of view, the axioms are valid in that they are a translation of the necessary and sufficient conditions for coherence (our starting point in Chapters 3 and 4). It follows that no other axioms can be admitted (since these would introduce further restrictions).

Mention should be made of a formulation which is subjectivistic in a purely psychological sense, and in which no axioms would be acceptable. This is the approach in which one simply thinks of evaluations of probability – in general, incoherent – being made by some, arbitrary, individual. It is clear that without sufficient preparation and thought everyone would give incoherent answers in every field (e.g. by estimating distances, areas, speeds etc. in an incoherent manner). This does not imply, however, that there exists – albeit only in the individual's own mind – a different theory (e.g. a non-Euclidean geometry) to be made an object of study. The object of study could only be the extent of his intuitive inability to understand the conditions of coherence, and to avoid breaking them. Otherwise, one would have to say that, in a system of bets, he deliberately chooses to behave in such a way as to lose.

From the classical point of view – probability 'defined' as the ratio of favourable cases to possible cases, all considered 'objectively' equally likely for reasons of symmetry – the axioms are true by virtue of the laws of arithmetic (sums of fractions, together with certain other details which are required to achieve the necessary rigour). There is, for any given application, just one admissible **P**. It would appear to be valid to consider infinite partitions into equally probable cases (by virtue of symmetry).

As an extension of this point of view, one might consider the 'necessary' conception, which takes probabilities of a collection of events, possibly outside the range of cases considered in the 'classical' approach, to be uniquely defined for *logical* reasons. A typical example – one that accepts the possibility of 'an infinite number of equally likely cases' – is provided by Jeffreys' admission of improper initial distributions (e.g. uniform in X, or in $\log X$ etc.). It appears that Carnap's point of view is similar to this.[52]

From the statistical point of view – where probability is regarded as 'idealized frequency'[53] – additivity always holds for arithmetic reasons (as in the classical case). According to this conception (again as in the classical case), there should be a unique admissible **P**. It is difficult to attempt to venture hypotheses about the interpretation of more delicate cases (e.g. zero probability).

An attempt to make the statistical conception more precise consists in defining probability not as an 'idealization', but rather as a *limit* of the frequency (as the number of trials, n, tends to ∞). In throwing a die, the limit-frequency of 'evens' is without doubt the sum of the limit-frequencies of '2', '4' and '6', if these limits exist (and this is assumed

52 It is always difficult to judge whether similarities are real or apparent (particularly between authors with different backgrounds, working in different fields).

53 This phrase does not really convey anything, but it is the only way to refer to the many confused explanations given by the supporters of this conception, and it may be that, in fact, there is nothing of substance to 'understand' (alternatively, it may be me who lacks the resources necessary for success in this toilsome venture).

to be the case in the scheme we are considering). It seems equally clear that such additivity does not (necessarily) hold for infinite partitions: a 'die with an infinite (countable) number of faces' could well turn up each face with limit-frequency zero[54] (indeed, *it should* do so, if we continue to admit, in some shape or form, the assumption of equal probabilities for all the infinite faces). But this seems to get overlooked (see Chapter 3, 3.11.6, the case $C9 = N9$).

Although it is really going beyond our aim of giving a critical analysis of particular attempts at axiomatization, it seems necessary to spend some time on the following case in order to make some comments (we shall refer to the version given by von Mises, which is the most developed, and which was in favour for a time).

On the one hand, it appears that the hypothesized sequences with well-determined limit-frequencies should represent 'idealizations' of problems of 'repeated trials'. This would seem to be so in view of introductory remarks alluding to ideas like the 'empirical law of averages' and because of an additional restriction ('Regellosigkeitsaxiom', or the axiom of 'nonregularity'), which is intended as a summary, in objective and descriptive terms, of the apparent effects of the *independence* of successive trials. Should one exclude periodicities? The grouping of the results in blocks (e.g. each 'colour' at least three times in a row)? The sequences definable in terms of simple mathematical formulae, or sentences not exceeding 100 words? On each occasion one would probably answer yes; but in actual fact there is never any reason to call a halt before having excluded all the possibilities, nor, conversely, any justification for absolutely excluding any given case.

On the other hand, if one wishes to consider the actual case of a sequence of trials (in general unlimited, but to begin with let us assume it to be limited to a finite number, n), independent and with equal probabilities – according to the concepts derived from such a formulation – one must not think in terms of the 'sequences' of the previous model. One must think of n parallel sequences; that is a sequence of n-tuples representing (let us say) a fictitious infinity of 'copies' of the actual sequence of n trials (for $n = \infty$, a fictitious infinity of copies of the whole actual sequence). Only in this absurd supermodel do the simple and obvious ideas of independence and equal probabilities make any sense, and is one able to (correctly) conclude that in the actual (Bernoulli) sequence one has stochastic convergence (weak, strong, in mean-square), but not definite convergence (as was postulated in the original scheme).

---

54 A 'reasonable enough' example might be obtained by saying that each face of the die has probability $p_h = e^{-1}|h!$ of occurring h times (in an infinite series of trials). Alternatively, if one prefers, one could say that out of 10 000 faces occurring 'at random' we will have 'in prevision' (or, less appropriately, 'on average')

$10\,000\, p_h =$	3679	3679	1839	613	153	31	5	1
where $h =$	0	1	2	3	4	5	6	7

(i.e., in the infinite series of trials 0, 1, 2,…, 7 repeats will occur, but 8 or more will not occur even once – on average about 0·10 times).

This is the Poisson distribution with $m = 1$, which holds asymptotically for the game of matching n objects ($n \to \infty$), or for that of n drawings with n balls (e.g. 90 drawings, with replacement, of the 90 numbers in bingo), again as $n \to \infty$. We have remarked that this example is 'reasonable enough', but it is no more than this, because the choice of this scheme from among infinitely many others is arbitrary. One could, for example, vary the scheme for drawing the balls, by assuming that as n, the number of balls, increases, the number of drawings is not n, but $2n$, or n, or $\sqrt{n}$ etc.

The original scheme is therefore a sham. For the purpose of winning over the unwary – who do not notice the sleight of hand – properties are attributed to it which look like the probable properties of actual sequences of 'repeated trials', but which are, in fact, misleading and incompatible with them. By mysterious manipulations of an infinite number of such shams, one finally succeeds in saying those things which could have been said directly anyway (i.e. that the trials are independent and equally probable). The fruits of these labours are that one now does not understand the (subjective) meaning of the words, and that the flurry of sophistical acrobatics has created the illusion that one has established or produced an 'objective' something or other.

In bringing to a close our summary of the various interpretations, we repeat that the subjectivistic conception is not in opposition to any of them, but rather that it utilizes all of them. It is simply a question of rejecting the claims of exclusiveness that lead to incomplete and one-sided theories, of correcting the distortions made in order to make them appear objectivistic, of considering them as methods whose appropriateness varies with the situation, and of seeing them as having one and the same function: that of aiding the individual in his task of evaluating the probabilities (always subjective) to be attributed to events of interest.

18 Questions Concerning the Mathematical Aspects

18.1. We now turn to an examination of those aspects of a purely mathematical or formal nature. In a certain sense, we will be considering the properties of the function **P** and the meaning of the implications of these properties. We refer here to meaning in a formal sense; without reference – except incidentally, and for purposes of clarification – to the different interpretations and assumptions which precede the choice of axioms (concerning which see Section 17).

In order to provide an overall perspective, it will be convenient to present the various questions – including those we have already dealt with – in the context of a comparison with the Kolmogorov axiom system,[55] a formulation which is well known to everyone.

The basic differences are:

i) we REJECT the idea of 'atomic events', and hence the systematic interpretation of events as sets; we REJECT a ready-made field of events (a Procrustean bed!), which imposes constraints on us; we REJECT any kind of restriction (such as, for example, that the events one can consider at some given moment, or in some given problem, should form a field);
ii) we REJECT the idea of a unique **P**, attached once and for all to the field of events under consideration; instead, one should characterize *all* the admissible **P** (the set P); P turns out to be closed (Chapter 3, Section 3.13), and thus any **P** adherent to P in fact belongs to it (a property which does not hold in the Kolmogorov system);

55 A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin (1933). The first time I developed a systematic discussion in the context of a comparison with this theory was in 'Sull'impostazione assiomatica del calcolo delle probabilità', *Annali Triestini*, **XIX**, University of Trieste (1949).

iii) our approach deals directly with *random quantities* and *linear* operations upon them (events being included as a special case); we thus avoid the complications which arise when one deals with the less convenient Boolean operations;

iv) we REJECT countable additivity (i.e. σ-additivity);[56]

v) we REJECT the transformation of the theorem of compound probabilities into a definition of conditional probability, and we also REJECT the latter being made conditional on the assumption that $\mathbf{P}(H) \neq 0$; by virtue of the exclusions we have made in (iv) and (v), the construction of a complete theory of zero probability becomes possible;

vi) Kolmogorov's proof of the compatibility of his axioms is open to criticism (see the paper of mine quoted in the footnote at the beginning of this section); this is, however, a problem that can be resolved, and it has no substantive implications.

To a greater or lesser extent, all these matters have been touched upon already, either in the text or in this Appendix. We shall only concern ourselves now with those aspects which require further analysis or more detailed discussion.

18.2. *Zero probabilities*. Let us first of all go back to our earlier discussion (at the end of Section 16), and let us repeat the proofs and definitions that we gave (in Chapter 4, 4.18.2), basing ourselves now on Axiom 3.

Axiom 3 permits us to define the ratio of the probabilities of two arbitrary events, A and B, by observing that for all H which contain A and B (i.e. $H \supset A \vee B$) the ratio $\mathbf{P}(A|H)/\mathbf{P}(B|H)$ does not change (except possibly to become indeterminate, 0/0). Suppose, in fact, that H' and H'' are events containing $A \vee B$, and that they do not give rise to the case of 0/0, and let $H = H'H''$ be their product, which also contains A and B (or one could take H to be $A \vee B$). Since $\mathbf{P}_{H'}$, and $\mathbf{P}_{H''}$ must be coherent, we can write

$$\mathbf{P}_{H'}(A) = \mathbf{P}_{H'}(AH) = \mathbf{P}_{H'}(H).\mathbf{P}_{H'}(A|H).$$

But $\mathbf{P}_{H'}(A|H) = \mathbf{P}(A|HH') = \mathbf{P}(A|H)$, because $H \subset H'$, $HH' = H$. Finally, we have $\mathbf{P}_{H'}(A) = \mathbf{P}_{H'}(H).\mathbf{P}(\nu)$, and $\mathbf{P}_{H'}(B) = \mathbf{P}_{H'}(H).\mathbf{P}(B|H)$, and hence it follows that

$$\frac{\mathbf{P}_{H'}(A)}{\mathbf{P}_{H'}(B)} = \frac{\mathbf{P}_{H'}(H).\mathbf{P}(A|H)}{\mathbf{P}_{H'}(H).\mathbf{P}(B|H)} = \frac{\mathbf{P}(A|H)}{\mathbf{P}(B|H)}.$$

The same holds true for every H'', and, finally, in order to obtain the ratio, it suffices simply to take $H = A \vee B = A + B - AB$; in this case we certainly have $\mathbf{P}_H(A) + \mathbf{P}_H(B) \geqslant 1$ and 0/0 cannot occur.

In this way, the formula $\mathbf{P}(E|H) = \mathbf{P}(EH)/\mathbf{P}(H)$ *is always meaningful and valid*, and the same is true for every application of the theorem of compound probabilities, and, more generally, for any operation involving probability ratios, so long as they make

56 It is worth mentioning, incidentally, that if one decides to proceed in the direction of assuming things for 'mathematical convenience', then not even countable additivity appears to be sufficiently restrictive. Several authors, including Kolmogorov himself, have recently proposed axioms ('perfect' additivity, and such-like) that make the principles of probabilistic reasoning, essential to every human being, completely dependent on the abstruse subtleties of set theory at its most profound. See, for example, D. Blackwell, 'On a class of probability spaces', in *Proc. 3rd Berkeley Symp.*, II, pp. 1–6 (1956), and other works referred to therein.

sense (i.e. so long as one does not introduce the nonsensical, indeterminate expressions, $0/0$, $0/\infty$, ∞/∞, which must be avoided by means of the procedure used for defining the ratio in each case).

There are, therefore, different orders, or layers, of zero probability (as we have already noted in Chapter 4, 4.18.3, where we also saw how very rich and complicated structures of such layers could be constructed). We shall see later, in 18.3, what the situation would necessarily be in this respect were we to assume the axiom of countable additivity (or, at least, if the condition were assumed to hold in some specific example or other). For the present, however, let us return to the general case.

The theorem of total probability will have to be interpreted in the following extended sense (which includes the case of zero probabilities): given n incompatible events, E_1, $E_2,\dots, E_n$, the probability of the sum-event E is the sum of the nonzero probabilities, if there exist any, and, if not, it is *the sum of the zero probabilities of maximal order*. If, for example, E_3 is of maximal order (i.e. $\mathbf{P}(E_h)/\mathbf{P}(E_3) < \infty$, $h = 1, 2,\dots, n$), the sum-event has probability $\mathbf{P}(E) = \mathbf{P}(E_3)$ if for all $h \neq 3$ the preceding ratio is not only $<\infty$ but in fact $=0$. In general, it is given by

$$\mathbf{P}(E) = \mathbf{P}(E_3)\sum_{h=1}^{n} c_h,$$

where $c_h = \mathbf{P}(E_h)/\mathbf{P}(E_3)$ ($c_3 = 1$; the other c_h may be zero, in which case they do not count, or they may be greater than, or less than, 1).

The introduction of conditional probability freed from the restriction that the 'hypothesis' have nonzero probability, and the consequent possibility of comparing zero probabilities, is important both from a conceptual and from a practical point of view. This importance derives not so much from the fact that we can see the potential usefulness in interesting applications, but rather from the warning it provides against inaccurate ways of approaching – or, at least, of expressing – certain questions. We have in mind methods of approach that either lead to confusion or to an over-hasty choice of the path and interpretation to be followed, because of the absence of a precise meaning to which one can refer.

No doubt some will regard discussion of this kind as rather artificial and academic; nothing more than hair-splitting *ad infinitum*. They may be right and they will do well to pose their problems in such a way as to avoid the difficulties. But in order to do this, they must first be able to recognize the difficulties as such, so as to overcome them without lapsing into naïvety or contradiction. In any case, since there do exist differences of opinion in this respect, and since the one which I consider to be correct, and which I uphold, differs from that which forms an integral part of the theory currently most in favour, there is no alternative, in the present context, but to consider the matter more deeply.

But why worry about events with zero probability? Are they not, for this very reason, eventualities which can be ignored?

From time to time, someone imagines that he had discovered the way of eliminating the problem altogether, by *establishing* that the values 0 and 1 must be reserved for the probabilities of the impossible event and the certain event, respectively. Every possible event should have positive probability (strictly less than 1). It is easy to see – and we shall do so presently – that this leads in our case to the same kind of absurdities as one encounters when trying to invent a measure which only assigns zero to the empty set. It is only in the very simplest examples (i.e. those where we only meet finite or countable partitions) that it may happen, *by chance*, that there are no possible events with zero probability, or that, if there are an infinite number of them, their union still has zero probability (a case in which, to some extent, one might regard them as eventualities that can be ignored). If, on the other hand, we considered a nondenumerable partition, we would have to conclude that it was impossible to consider a nondenumerable partition into *possible* cases (because at most a countable number can have positive probability if their sum is not to become infinite $\sum p_h = \infty$). We need go no further than the logic of certainty to see the absurdity of this statement.

The major difference between events of zero probability and impossible events is the following: the union of an infinite number of the former can have a nonzero probability (and may even be the certain event), whereas any union of the latter can only be an impossible event.

It is in this setting that one comes across the most controversial question of them all; that of 'countable additivity'. If we limit ourselves to a discussion of it in the context of events having zero probability, countable additivity implies that taking a countable union will never yield an event with positive probability (and certainly not the certain event). One has to examine the specific question of whether it is possible and appropriate to assume this property as an axiom of probability (the property holds, as is well known, for Lebesgue measure, where the nature of the definition excludes the cases for which it would not hold[57]). The majority opinion is that the answer is yes. In my opinion, this is a consequence of external factors, which, generally speaking, are not examined in order to check whether or not they correspond to the essential nature of the problem.

Certain aspects of conditional events also involve us in a consideration of the problems that derive from the presence of events with zero probability. If an event is possible, then – independently of the probability attributed to it, even if it is zero – events conditional on it, and bets related to these events, can always be considered. In this way (by means of the above-mentioned formulae, which we need not consider here), it becomes possible to compare all zero probabilities. They may be of the same order (i.e. having a finite ratio); or of a different order (ratio equal to zero, or, conversely, to infinity). For the purpose of providing an analogy, we have a situation similar to that which arises in comparing two geometrical objects of zero volume; they can be compared by considering the ratios of their areas, or of their lengths, depending on whether they are both two dimensional, or one dimensional. On the other hand, we would say that one was of smaller order if it were a line segment while the other were part of a plane.[58]

We are not concerned with pushing the analogy too far, because the geometrical case has certain special features of its own. What is common to both is the idea of measures of different orders (or, if one prefers, of non-Archimedean quantities). However, the example is to be understood in a purely illustrative sense, with a warning that one should not take into account notions like dimension, distance, volume, limit, cardinality and so on.

---

57 As in the example given by Vitali (quoted in Chapter 6, 6.5.9). See G. Vitali and G. Sansone, *Moderna teoria delle funzioni di variabile reale*, Zanichelli, Bologna (1935), part I, pp. 56 ff.

58 A more systematic method of comparison – for simplicity, we shall always refer to ordinary, three-dimensional space – would be to consider, for each set I, the set of points I_ρ whose distance from I is less than ρ, and the function $V_I(\rho) = \mathrm{Vol}(I_\rho)$ (volume of I_ρ). One can now define the *ratio of the measure of two sets* I' and I'' to be the limit as $\rho \to 0$ of the ratio $V_I'(\rho)/V_I''(\rho)$ (if it exists). It does not always exist, but in the most 'regular' cases we have $V(\rho) = k\rho^{3-d}(1 + 0(\rho))$; in other words, $V(\rho)$ is comparable with a power ρ^α $(0 \leqslant \alpha \leqslant 3)$, and, in particular, volumes, areas, lengths, the number of isolated points are given by the coefficients k in the cases for which α turns out to be 0, 1, 2, 3 ($d = 3 - \alpha$ is the number of dimensions, $d = 3, 2, 1, 0$).

Let us just mention that the consideration of probability as a non-Archimedean quantity would permit us to say, if we wished, that 'zero probabilities' are in fact 'infinitely small' (actual infinitesimals) and only that of the impossible event is *zero*. Nothing is really altered by this change in terminology but it might sometimes be useful as a way of overcoming preconceived ideas. It has been said that to assume that

$$0+0+0+\ldots+0+\ldots=1$$

is absurd, whereas, if at all, this would be true if 'actual infinitesimal' were substituted in place of 'zero'. There is nothing to prevent one from expressing things in this way, apart from the fact that it is a useless complication of language, and leads one to puzzle over 'les infiniment petits' [the infinitely small].

Despite all that has been said, some readers may still be of the opinion that all these things are pointless hair-splitting anyway – and, in a certain sense, I would like to reply YES. The fact remains, however, that, paradoxically, the only way of dealing with these things is to think about them and analyse them in detail, carefully studying the most valid and appropriate way of setting them aside, case by case. Even in those cases where approximate answers are preferable to exact ones (because of the illusory nature of the exactness), it is especially important to be doubly precise in one's arguments, in order to know which things remain valid, and which require modification, when the reasons for, and the degree of, this illusory exactness are taken into account.

To this end, we shall make use, among other things, of the ideas that were developed concerning the 'precision' factor, and we shall arrive at conclusions which (hopefully) will appear reasonable, sensible and, perhaps, obvious. But this feeling will only be justified when we have arrived at the conclusion by means of an accurate evaluation of alternative suggestions, which clarifies just what is, and what is not, really significant and well founded.

18.3. *Countable additivity*. An extensive treatment of this topic was given in Chapter 3, Section 3.11, and we have also referred to it on many subsequent occasions. Let us recall the main points as a prelude to making some further critical comments.

The property of additivity, which we have assumed as an axiom, says that in a *finite partition* the sum of the probabilities must equal 1. In other words, if $E_1, E_2, \ldots, E_n$ are exclusive and exhaustive, the probabilities $p_1, p_2, \ldots, p_n$ attributed to them must be non-negative with sum equal to 1. In fact, this is not merely a necessary condition for the evaluation to be coherent and admissible but it is also sufficient.

In the case of an infinite partition into events E_h ($h \in H$, where H is arbitrary), we can only say, on the basis of our axiom, that the sum of every finite number of the p_h must be $\leqslant 1$: in other words, that at most a countable number of them can be positive ($\neq 0$), and that for such values[59] we must have $\sum p_h \leqslant 1$. If, in particular, the set of positive p_h has sum = 1, then the E_h with zero probability also have zero probability when taken

---

59 Even if there are an infinite number of them one can speak of a 'sum' in the sense of 'upper bound of the sum of a finite number of terms' (if one thinks of the 'sum of the series' no conclusion would be legitimate). If we denote by Σ the upper bound (possibly $+\infty$) of the sums of a finite number of terms, we may denote in this way the sum of an arbitrary infinite number of non-negative numbers, and, in particular, of events. For example, $\Sigma E_h (h \in k)$ will denote the number of successes among those E_h for which $h \in K$, and we observe that the standard convention – see Chapter 1, Section 1.9 – in which $(h \in K) = 1$ or $(h \in K) = 0$ according as h belongs, or does not belong, to K, allows one to write $(h \in K)$ as a factor, on the same line, instead of as an index written below the Σ sign (see the application, which follows shortly, with $(p_h = 0)$ for $(h \in K)$ where K = 'the set of the indices for which $p_h = 0$'). If the E_h are incompatible, the sum is necessarily either 0 or 1.

together: that is it turns out that the union $E = \sum E_h$ ($p_h = 0$) also has zero probability, $\mathbf{P}(E) = 0$. If, on the other hand, we obtain $\sum p_h = P < 1$, that is if a probability $1 - P$ is *missing* in the partition, then the possibilities are as follows: if there are a finite number of events with nonzero probability, then this missing probability is necessarily that of E = the union of the events with zero probability; otherwise, it may be attributed arbitrarily to E and to $\tilde{E}$[46] = the union of the events with positive probability;

$$\mathbf{P}(E) = P', \quad \mathbf{P}(\tilde{E}) = P + P'', \quad P' + P'' = 1 - P.$$

To summarize: given the probabilities p_h of the events E_h of a partition, if their sum is = 1 then the probabilities of all the events depending on it – that is sums of a finite or infinite number of events in the partition – are uniquely determined. In this case, we shall say that the probability $\mathbf{P}(E)$ is countably additive on the partition $\{E_h\}$. Otherwise, this only holds for event-sums of a finite number of the E_h, or for their complements: in any other case, a margin of indeterminacy equal to $1 - P = 1 - \sum p_h$ remains;

$$p' = \sum p_h \left(E_h \subset E\right) \leqslant \mathbf{P}(E) \leqslant 1 - \sum p_h \left(E_h \subset \tilde{E}\right) = p'' = p' + (1 - P).$$

Let us be clear that 'indeterminacy' simply means that the extension is not, in general, uniquely defined; one only has bounds. There is no 'indeterminacy' in any specific sense; such as being 'barred' from attributing a well-defined value to $\mathbf{P}(E)$. It is simply that E is not one of the events whose probability has already implicitly been evaluated by virtue of our evaluations for the E_h; it is just one of the many for which our choice is more or less open. We are completely free in our choice (i.e. can give $\mathbf{P}(E)$ any value between 0 and 1) in the particular case in which all the events of the partition have been attributed zero probability (and E is not the sum of a finite number of the E_h, nor the complement of such a sum). This happens in the case of a continuous distribution (on the line, or in the plane, or in ordinary space, …) for which we have established only that it is 'without concentrated masses', or for a countable number of exclusive (and exhaustive) events of zero probability.

Conclusions of this kind may be hard to accept, or perhaps may even appear paradoxical. At least, the way in which many authors bend over backwards to avoid them – by introducing some new axiom (or 'strengthening' the existing ones) – seems to suggest that this is the case. The following are some of the kinds of restrictions which could be imposed:

(Z) denying that it is legitimate to attribute zero probability to a possible event;
(Za) denying that a union of events with zero probability can have nonzero probability;
(Zb) as (Za), but only considering a countable number of events;
(Ka) assuming countable additivity for arbitrary partitions;
(Kb) as (Ka), but only for countable partitions.

We have introduced the letters Z, Za, Zb, Ka, Kb, in order to facilitate references to these 'axioms'; in what follows, we shall, of course, argue against them.

The first and the last have actually been proposed; Za and Zb are progressively weaker versions of Z and are also special cases of Ka, Kb, respectively; the inclusion of intermediate possibilities would only serve the purpose of pointing out these connections.

First of all, we should draw attention to the lack of any real arguments on the part of those who support such a restriction. It is usually presented as a 'natural' extension of the theorem of total probability as $n \to \infty$ (as, for example, in Cramèr); or as a 'natural' property by analogy with Lebesgue measure (and this is the most common idea); or by Baire extension by continuity (like, for example, in Feller). In other words, for 'mathematical' reasons, and not for reasons relating to probability theory.

One mathematical consequence of this is that it becomes impossible to think of a $\mathbf{P}(E)$ defined for all the events which could be formed on the basis of a nondenumerable partition. An example is provided by the power set of any set whose cardinality is that of the continuum (by virtue of the results of Vitali, Lebesgue, Banach, Kuratowski and Ulam, concerning the impossibility – except in trivial cases of 'concentrated mass' at a finite or countable number of points – of extending σ-additive measures to all the subsets of a nondenumerable set[60]). To admit σ-additivity is to contradict the basic idea that one can attribute to any uncertain event whatsoever a probability – without any, logically inexplicable, discrimination between one event and another. Of course, it could happen that this 'basic idea' itself gives rise to conflicts with other requirements: for example, were it true that there does not always exist an extension of a finitely additive $\mathbf{P}$, we should have had to re-examine the whole question of whether, and in what way, a mathematical theory of probability was possible (with goodness knows what weakening of the axioms[61]). The fact that such a disaster does not occur for finite additivity, but does occur if one attempts to replace it with σ-additivity, clearly indicates that the substitution is entirely inappropriate.

If one accepts the subjective concept of probability, the conclusion becomes even more obvious.

In order to reach this conclusion, it was not even necessary, in fact, that the contradiction of not being able to find an arbitrary $\mathbf{P}$ came to light. It was sufficient that the choice was restricted in a way which appeared to preclude each individual being permitted an unfettered evaluation. And this occurs even in the case of a countable partition. Let us suppose that there are a countable infinity of 'possible cases' and – in order to avoid thinking of points or sets on the real line which appear 'special' in some way – let us imagine that they are represented by points on the circumference of a circle, whose distance apart is a rational multiple of 2π (i.e. by taking the origin at an arbitrary one of these points, with $\theta = 2k\pi$, k rational, $0 \leqslant k \leqslant 1$). That an axiom should not permit me to attribute probabilities which are negative, or have sum greater than one, is something which can be clearly understood as a condition of coherence; it does not impose any

60 From our point of view, it suffices that this has been established for some sets. It appeared preferable, therefore, not to weigh down the text with details of how the result has been proved (Ulam) provided that the cardinality of the set is not 'inaccessible'.

61 A more 'minor' difficulty may serve as an example. A paradox, due to Haussdorff, says that a spherical surface can be divided into three sets, A, B, C such that each is superimposable both on each of the others and on their union; it follows that any 'measure' which was finitely additive and invariant under rigid motions would assign to these sets both $\frac{1}{3}$ and $\frac{1}{2}$ (and $\frac{2}{3}$ and so on; as is well-known, one can logically deduce anything if one starts from something ridiculous). This contradicts geometric intuition, but not the idea of probability (nor the 'axiom of choice'). As Paul Lévy said, in order to refute this interpretation, the simple fact is that 'the continuous in higher dimensions is even more complicated than we thought'. (See E. Borel, *Les paradoxes de l'infini*, Gallimard, Paris (1946); Paul Lévy, 'Les paradoxes de l'infini et le calcul des probabilités', and a note by Borel, in *Bull. Sci. Math* (1948), pp. 184–192.)

restriction on my freedom of opinion. But suppose that an axiom (like *Zb* or *Kb*) prohibits me from attributing the same probability, $p_h = 0$, to all the events; or even (like *Kb*) forces me to choose some finite subset of them to which I attribute a total probability of at least 99% (leaving 1% for the remainder; and I could have said 99·999% with 0·001% remaining, or something even more extreme). If I do not happen to hold these opinions, and have no reasons for adopting them, then this is no longer a question of coherence; it is a direct interference with my judgement![62]

Moreover, to permit the assignment of zero probabilities to all the events (of a countable partition) is a much less restrictive idea than, as in the finite case, considering them as 'equally probable' ($\mathbf{P}(E_h) = 1/n$, $h = 1, 2, \ldots, n$). The equivalent of this would be to consider the $\mathbf{P}(E_h)$ equal, not in the sense that $\mathbf{P}(E_h) = \mathbf{P}(E_k) = 0$ as real numbers, but in the sense that $\mathbf{P}(E_n)/\mathbf{P}(E_k) = 1$ (as a ratio of zero, or 'infinitely small', probabilities). For the first condition ($p_h = 0$) to hold, much less is required: in terms of ratios, it is sufficient that there do not exist probabilities of maximal order (for example, that give, for each h, $\mathbf{P}(E_{h+1})/\mathbf{P}(E_h) = \infty$), or that, if they do exist, their sum (taking one of them to be unity) is infinite. (It is also sufficient – but this cannot be derived from the ratios alone – that the probability of all the cases be infinite in the given scale; and it may happen that this occurs for the union of cases with probability of smaller order without occurring for those of maximal order.)

Assuming (in line with Axiom 3) that, if accepted, axioms *Zb* or *Kb* should also hold for probabilities conditional on an arbitrary possible event H, they would imply even more restrictive conditions for the probabilities of individual 'possible cases', in order to avoid – pulling out a countable number of them – the possibility of a case of probabilities all zero. There could be at most a countable number with the same order, so that, given a non-denumerable infinity of 'possible cases', we should have a nondenumerable infinity of different 'orders' of probability.[63]

We should also mention that, from time to time, problems which, explicitly or implicitly, run counter to the assumption of countable additivity are also considered by authors who insist on the latter as an axiom. Sometimes the case of 'an integer chosen at random' is regarded as 'meaningful' but 'breaking the rules' (the probability taken to be the limit density; for example, the probability that the integer is a multiple of $k = \lim$ [(number of multiples of k between 1 and $n)/n] = 1/k$). At other times (see, for example, Rényi, Chapter 3, 3.18.5), one considers conditional probabilities; for instance, a distribution inside a circle, which is then made larger and larger, so that the probability of each finite region tends to zero. Countable additivity is prescribed for the conditional probability (i.e. within any circle), but it is not made clear that this no longer holds for the limit distribution (which is not explicitly dealt with in its own right, the passage to the limit being merely a device).

This seems to provide further evidence in support of our initial impression that the assumption of countable additivity owes very little to genuine probabilistic

62 It is strange that the very same people who, in general, would encourage one in the finite case to accept a judgement of equal probabilities, on the grounds that a person 'knows nothing', seek to prohibit someone who, on the grounds that he 'knows nothing', would like to make the same judgement in the countably infinite case.

63 One can say even more: those of the same order must possess a convergent sum; the different 'orders', arranged in decreasing order, must form a 'well-ordered' sequence, so that there always exists a 'maximal order'.

considerations; in other words, that it is more a mathematical embellishment than a necessary property of probability. Many other anomalies and peculiarities (one might even say – in a psychological rather than a formal sense – contradictions) strengthen the same impression. The fact that 'equal probabilities' are perfectly acceptable in the finite and continuous (points of an interval) cases, but are not allowed in the countable case, can be explained only by drawing attention to our habit of applying, in particular cases, the most widely used tools (in the countable case we are accustomed to summing series!), rather than adhering to the principle of coherence. The very fact that one treats the finite and uncountable cases differently from the countable case (axioms *Zb* and *Kb*) is sufficient to show that more thought is given to the mathematical structure than to the logical problem – for which the meaningful distinction, if any, would seem to be that between the finite and the infinite (of whatever kind).

Bearing this in mind, the line we have followed here seems to represent, independently of the reasoning we have put forward – which, hopefully, is more persuasive – the most natural way of connecting together attitudes that are, at least in part, inspired by fragmentary and irreconcilable points of view.

Finally, we should mention a concept and a result that have come to be considered as a justification for the systematic use of σ-additivity. The basic idea is the possibility of stretching the interpretation in such a way as to be able to attribute the 'missing' probability in the partition to new fictitious entities in order that everything adds up properly. In some cases, in order to salvage countable additivity, it is even claimed that the new entities are not fictitious, but real. I remember having seen something of this kind in a paper (by Kingman, I think) which involved a probability distribution for discrete processes concentrated in the neighbourhood of a limit case where the process would become continuous.[64] This was taken as an indication of the necessity of including the continuous limit cases among the possible cases, in order to be able to foist upon them the probability missing from the sum.

Using this kind of argument, one could say that if the possible cases were the rationals, and if to each of them is attributed zero probability, then we have demonstrated that the real numbers must also *exist*, and be possible, because they are required as the indispensable support for the probability as a whole (=1).

The more general kinds of considerations we have alluded to are more abstract, although, in the final analysis, they reduce to the same type of argument as is involved in the addition of fictitious entities. In mathematics, this kind of argument or procedure is well known to be fruitful (like, for example, the addition of new points in order to compact a space), but, in our case, events must be events, and not abstractions, if the theory is to preserve a concrete meaning; that is to say 'has some meaning', and this, in the formulations we have mentioned, does not happen.

In fact, it is necessary to have recourse to 'ultrafilters' (and this gives only a theoretical possibility of obtaining the desired result). In any case, this would only hold in a field that has been modified with respect to the original one, and the latter is the only one that we are interested in. I have never seen any application to the study of actual cases (and it seems impossible that it should constitute a simplification, rather than an

64 *Translators' note.* J.F.C. Kingman, 'Additive set functions and the theory of probability', *Proc. Cam. Phil. Soc.*, 63 (1967), pp. 767–775.

unnecessary complication, introduced for the purpose of permitting yet another unnecessary complication, i.e. σ-additivity). It seems to me that the only result has been to encourage people even further to consider just those cases where σ-additivity holds directly, and to ignore the others because they can, in theory, be transformed in such a way as to turn out to enjoy, fictitiously, the property which, in the field one ought to be considering, does not actually hold.

In practice, there are quite different purposes for which consideration of ultrafilters can be useful. In particular, for studying '*agglutinated* probabilities' (i.e. probabilities that cannot be subdivided), which can arise in distributions.[65] Think of the case (which we have already mentioned) of distributions on an ultrafilter: an ultrafilter is a family of events (sets) to which one and only one element of a partition can belong, and we attribute probability 1 to the events belonging to it (and, therefore, probability 0 to the others).

The consideration of *filters* can also be useful if one wishes to analyse further the possibility of dividing up the *missing* probability. In general, given a partition into events E_h, with $\sum p_h < 1$, it is sufficient to consider another event B (or a partition, $B_1, B_2, \ldots, B_n$), and to form the partition BE_h (or the partitions $B_1E_h, B_2E_h, \ldots, B_nE_h$) of B (or of B_1, of $B_2, \ldots$, of B_n). The missing probability $1 - \sum p_h$ can then be divided up between the filters generated by B and by $\tilde{B}$(or by $B_1, B_2, \ldots, B_n$). Think, in particular, of the mass adherent to a point (in a distribution on the real line).

The probability adherent to the left (or to the right) may be further divided up by considering filters; for example, in the case of the sequence of sets I_n of rationals between $x - 1/n$ and x, one obtains the probability adherent from the left on the rationals, or on the irrationals and so on. Of course, just as knowledge of $F(x)$ is not sufficient to separate possible adherent masses from the concentrated ones, it is even less sufficient for these subdivisions, which have to be established on the basis of other considerations.

18.4. *Concerning what is 'reasonable'.* It would be very difficult to reach any conclusion or to make any constructive progress by attempting to conduct a discussion of this topic with supporters of opposing points of view. Each would first of all attempt to challenge the 'reasonableness' of the assumptions of the others, judging them to be too 'theoretical', lacking any concrete value, and based on the assumption of an absolutely unrealistic degree of precision.

There would be no difficulty for anyone in criticizing the formulations of others, and no doubt anyone making such criticisms against the formulation we have adopted here would find good reasons for so doing. The complications we have considered, however, do not arise, unless one wishes to isolate cases that are in a certain sense 'pathological'. For problems which are 'sensible from a practical point of view', one not only avoids these complications, but also those imposed everywhere *a priori* by the assumption of σ-additivity. The latter are harmful because they go beyond what is required in simple cases and, moreover, are over-restrictive in the complicated cases.

Our criticism of countable additivity on the grounds that it precludes one attributing probability to all events (for example, the extension of Jordan–Peano measure to all the

65 See B. de Finetti, 'La struttura delle distribuzioni in un insieme astratto qualsiasi', *Giorn. Ist. Ital. Attuari*, **XVIII** (1955), pp. 1–14. An English translation of this paper, 'The structure of distributions on abstract spaces', forms Chapter 7 of B. de Finetti, *Probability, Induction and Statistics*, John Wiley & Sons (1972).

sets of the interval [0,1]) is in no way intended as implying that the extension to Lebesgue measure is considered insufficient, or that one actually wishes to go further. On the contrary, it means that we consider it to be usually quite sufficient to confine attention to Jordan–Peano measure, but that, if one wishes to go further, the extension should be neither predetermined, nor ruled out in any way. In other words, Lebesgue measure is just one of the infinite number of extensions to larger families of sets and one should be free to choose any of these if one wishes to make such an extension. Should anyone opt for countable additivity as a matter of preference, there is no objection (just as, in a practical case, it is open to one to choose a distribution possessing a continuous density, rather than a less 'regular' one, without feeling that one is forced to make such a choice by virtue of some law of probability). Any other extension (to all sets) is equally legitimate (in principle: so far as its usefulness is concerned, it is not clear whether the Lebesgue extension should be regarded as useful, once it has been made clear that the consequences one derives from it are not consequences of the initial evaluation by virtue of the 'law' of countable additivity, but rather that they derive from the arbitrary choice of a particular one of the possible extensions of the given evaluation).

On the other hand, a similar criticism can be made at a much more basic and fundamental level. In many approaches, one establishes *a priori* that the probability $\mathbf{P}(E)$ has to be given for all the events E of some given family obeying some given conditions; for example, forming a field (in the above case a Borel field) that is considered fixed once and for all. One can then go on to consider only those problems that belong within that field (considered as a single, closed system, and often referred to as a 'probability space'). It is not then possible to evaluate the probabilities of two events A and B without doing the same for the product AB. But it could well be that sometimes one either has to, or wishes to, proceed (albeit temporarily) without the knowledge or evaluation of $\mathbf{P}(AB)$; or that the family of events initially considered (and 'arbitrary') does not contain all the products. The conclusions hold for all events and random quantities linearly dependent on those events one starts from. Indeed, we could start from random quantities, X, for which the previsions, $\mathbf{P}(X)$, were evaluated: the particular fact of whether all, or just a few, or none, of them are events is in itself irrelevant. The case of events seems simpler and more intuitive only because it is more familiar, as well as being more schematic, and capable of more varied representations (set-theoretic, for example). One may then examine for each problem (with no limitations of any kind) the implications of the evaluations already assumed made, and one can complete them by means of any further evaluations that are required to answer the questions of interest.

There is no need to make use of, or mention, probabilities conditional on events of zero probability (or to compare zero probabilities, as non-Archimedean quantities) except when this might be useful for more careful consideration of delicate situations – where it is otherwise easy to adopt a cursory attitude.

As is illustrated in the cases used as examples, cases which are representative of the general situation, the approach we adopt consists in keeping the treatment at the simplest and most concrete level, adhering to the practical meaning, rejecting assumptions that are not supported by compelling arguments (like that of replacing finite additivity by σ-additivity), rejecting the once-and-for-all fixing of closed structures and, instead, in always open-mindedly allowing the possibility of extending the probability field to be studied, as and when required.

In this sense, we obtain the maximum simplification. There are, however, certain circumstances in which complications do, in fact, arise. This happens when the assumption of σ-additivity leads to simple conclusions (well known from measure theory), which either no longer hold if we abandon the assumption, or require more careful and less 'intuitive' formulations. What should we reply to someone who objects to complications of this kind?

Our reply is that such complications are inherent in the fact that when we speak of probability (or prevision) we are referring to functions **P** which may well be finitely additive (instead of being assumed, *a priori*, by virtue of an 'axiom', to be σ-additive over the field under consideration; or, and it amounts to the same thing, that the term 'event' can only be applied to members of some σ-field, restricted in such a way that **P** is σ-additive over it).

There are, therefore, two possibilities. On the one hand, it may be that we wish to be able to make a statement that holds only under the assumption of σ-additivity; in this case, one can state it as it is, making explicit the assumption that **P** be σ-additive (over some given field, or, often, just over some particular partition). There are no complications, apart from that of stating the assumption, and this has the advantage over the abandoned axiom in that it only requires the assumption of the latter over the minimal field for which it is required. The approach is rather like forcing a person to declare that a property holds for continuous functions, or for functions continuous in a given interval, or at a particular point, when the person is accustomed to stating it as valid for all functions (leaving it to be understood that he is only referring to functions which are continuous everywhere).

Alternatively, it may be that one wishes the statement to be valid without the restrictive hypothesis under which those things that held under σ-additivity continue to be true. In this case, matters become more complicated (unless, for reasons of simplicity, one prefers an inaccurate statement). To put it concisely, there is always the possibility of choice: either stick to the assumption of σ-additivity (no longer considered as an axiom), making it clear that one is doing so, or state things in the form necessary for them to turn out to be true independently of this assumption.

It will suffice to recall various of the cases we have already examined (the reader can, if necessary, refer back to the extensive discussion given in the text); the possibility of masses adherent to a point (instead of concentrated at it) and, in particular, concentrated at infinity (improper distributions); the indeterminacy of $\mathbf{P}(X)$ with respect to distributional knowledge in the case of unbounded distributions; the bogus formulation of the 'strong law of large numbers' and related topics.

In many cases, simple, minor modifications of the kind put forward are sufficient to ensure the validity of a statement, independently of the axiom of countable additivity. Moreover – although this is a matter of taste – they serve, because of their 'finitistic' character, to give a more concrete air to things.

All previous considerations can be regarded as variations on a single fundamental theme: the desirability of basing oneself on axioms that are the weakest (i.e. least restrictive) from a mathematical point of view, because they are from a logical point of view the most securely based (i.e. the least disputable), and which lead to results and statements that are the most secure (i.e. the least disputable).

Let us consider again the introduction of $\hat{\mathbf{P}}(X)$ (Chapter 6, 6.5.7). The mathematical definition is unexceptionable and this, by the standards normally adopted, suffices to render $\hat{\mathbf{P}}(X)$ acceptable *by definition* as the value of $\mathbf{P}(X)$. In contrast, we put forward arguments whose purpose was to establish the existence of possible reasons for considering this choice as being, in addition, a 'reasonable' one (and this, in a certain sense, is even more important). We were careful, however, not to identify $\mathbf{P}(X)$ with $\hat{\mathbf{P}}(X)$. In fact, $\hat{\mathbf{P}}(X)$ is always just *one* of the possible values for the extension of $\mathbf{P}$ outside the field within which it is uniquely defined by the F (although, in a certain sense, it is the most 'reasonable' extension).

This example, and the discussion arising from it, serves also as an illustration of the kind of attitude that results from the choice of a 'conceptual' approach as opposed to a 'formal' one, in the sense already considered. All the ideas and results are drawn from the *meaning* that lies behind the axioms, and not from the mathematical conventions. In contrast to the tendency towards uniquely determining the extension of certain notions by means of special forms of passage to the limit, our effort consists in not admitting, even inadvertently, any restriction that is not the result of simple finitistic inequalities, and which – one might say – goes beyond the idea of the 'method of exhaustion'.

It is not a question of weighing up, *a priori*, one's preferences for this or that mathematical approach but, on the contrary, of emphasizing the need to choose, in any application, the tools most suited to the nature and meaning of the problem. The nature and meaning must not be distorted or disguised in order to introduce tools of a more or less elegant, sophisticated, or 'fashionable' kind.

18.5. *Countable additivity as continuity*. We return to the topic of countable additivity once again, this time in a different (although equivalent and suggestive) guise. We shall examine some further questions and provide further discussion.

The condition of countable additivity for events (as considered so far) can be expressed in an even more meaningful form as a 'continuity' condition. This is the condition which appears among the axioms given by Kolmogorov (and other authors) in the following form (which we shall call 'axiom' Kb'):

> *if $E_1, E_2, \ldots, E_n, \ldots$ is a sequence of events, each of which is contained in the preceding one, and whose product is empty (i.e. there are no 'elementary outcomes' common to all the E_n), then $\mathbf{P}(E_n) \to 0$ as $n \to \infty$.*

We can see immediately that this condition is equivalent to countable additivity. Let us write

$$E_1 = (E_1 - E_2) + (E_2 - E_3) + \ldots + (E_{n-1} - E_n) + E_n;$$

all the terms in brackets are events by virtue of the inclusion hypothesis, and the probability of E_1 is the sum of the probabilities of the $(E_h - E_{h+1})$ up to some point, plus the remainder. If the latter tends to zero, as axiom Kb' requires, the probability is given by the sum of the series, and countable additivity holds. The argument can be turned around straightforwardly: starting from a sequence $C_1, C_2, \ldots, C_n, \ldots$ of incompatible events, and setting $E_n = C_n + C_{n+1} + \ldots$, we reduce to the preceding case (with $C_n = (E_n - E_{n+1})$); in order that the series of the $\mathbf{P}(C_n)$ converges, the remainder, that is $\mathbf{P}(E_n)$, must tend to zero.

In fact, it is easily seen that Kb' leads, in general, to a further property, even more meaningful, and showing more clearly the appropriateness of the term 'continuity'. Note that for any sequence of events, or random quantities, we can consider the *lower limit* and the *upper limit* (and also, if these coincide, we can consider the *limit*, their common value), just as in analysis. The fact of whether the values of the sequence are known or not (random) is irrelevant. In particular, in the case of events, $E' = \lim\inf E_n$ and $E'' = \lim\sup E_n$ are the events that consist of the fact that a finite number of the E_n are *false* and an infinite number are *true*, respectively (i.e. a finite number take the value 0 and an infinite number take the value 1, respectively). To say that $E_n \to E$ (i.e. that E' and E'' *coincide*), or that the limit E of the sequence E_n exists, is to say that *necessarily*, in the case under consideration, from some N onwards the events E_n are either all true or all false (in other words, it is impossible for infinite sequences of both true and false events to occur).

Well, then: in the case of countable additivity one has

$$\mathbf{P}(E') = \mathbf{P}(\lim\inf E_n) \leqslant \lim\inf \mathbf{P}(E_n) \\ \leqslant \lim\sup \mathbf{P}(E_n) \leqslant \mathbf{P}(\lim\sup E_n) = \mathbf{P}(E'');$$

in particular, if the limit E exists, $\lim \mathbf{P}(E_n) = \mathbf{P}(\lim E_n) = \mathbf{P}(E)$.

It remains to check that the same condition holds more generally when we have a sequence of random quantities X_n rather than events.

The property

$$\mathbf{P}(X_n) \to 0 \quad \text{if } X_n \to 0$$

is valid (under the assumption of countable additivity) *if the random quantities X_n are uniformly bounded*. Suppose, in fact, that, for all n, $|X_n| < K$; then, for any (small) $\varepsilon > 0$, we have

$$|\mathbf{P}(X_n)| \leqslant \mathbf{P}(|X_n|) < \varepsilon + K\,.\,\mathbf{P}(|X_n| > \varepsilon)$$

(because $|X_n| < \varepsilon + K.(|X_n| > \varepsilon) = \varepsilon$ if $|X_n| < \varepsilon$, and $= \varepsilon + K$ otherwise). But if $X_n \to 0$ we also have $(|X_n| > \varepsilon) \to 0$, and this means – recall the above – that we cannot have an infinite number of $|X_n| > \varepsilon$, and hence (assuming Kb') $\mathbf{P}(|X_n| > \varepsilon) \to 0$. It follows that $\lim |\mathbf{P}(X_n)| < \varepsilon$, and, since ε is arbitrary, that

$$\lim \mathbf{P}(X_n) = \lim |\mathbf{P}(X_n)| = 0.$$

Since events are uniformly bounded ($|E_n| \leqslant 1$) the property we have established is equivalent to Kb' (i.e. to countable additivity). If we remove the condition of uniform boundedness, the property does not hold (even if countable additivity holds). Suppose we take a countable partition into events E_n to which we assign nonzero probabilities p_n with sum = 1, and let us consider the sequence of random quantities $X_n = E_n/p_n$ (which are not uniformly bounded). We obtain $\mathbf{P}(X_n) = p_n/p_n = 1 \to 1 \neq 0$, although $X_n \to 0$ (all the X_n but one are, in fact, = 0). By slightly modifying the example, putting $X_n = E_n / p_n^{\alpha}$ for instance, we obtain $\mathbf{P}(X_n) = p_n^{1-\alpha}$, and we see, therefore, that the property $\mathbf{P}(X_n) \to 0$ holds for $\alpha < 1$, whereas, if $\alpha > 1$, we have $\mathbf{P}(X_n) \to \infty$ (although, for the same reason as before, we still have $X_n \to 0$).

The extension of the property to the limit is similar. Assuming Kb, we have

$$\mathbf{P}(X') = \mathbf{P}(\lim\inf X_n) \leqslant \lim\inf \mathbf{P}(X_n) \\ \leqslant \lim\sup \mathbf{P}(X_n) \leqslant \mathbf{P}(\lim\sup X_n) = \mathbf{P}(X'')$$

if the X_n are uniformly bounded. We shall give the proof in the case of the upper limit (the other case is clearly symmetric) and our proof includes the case of events (where a proof was not given).

Putting $X''_n = \sup X_h$ (for $h \geqslant n$), $X'' = \lim\sup X_n = \inf X''_n$, we obtain $X''_n = X_n + (X''_n - X_n) = X'' + (X''_n - X'')$, where $(X''_n - X_n)$ and $(X''_n - X_n)$ are non-negative. We therefore have, in every case,

$$\mathbf{P}(X_n) \leqslant \mathbf{P}(X'') + \mathbf{P}(X''_n - X''),$$

and, if $\mathbf{P}(X''_n - X_n) \to 0$, we shall have lim sup $\mathbf{P}(X_n) \leqslant \mathbf{P}(X'')$. But

$$X''_n - X_n \to 0,$$

by definition, and, if we assume the X_n to be uniformly bounded (which implies, *a fortiori*, that the $X''_n - X_n$ are), then, assuming countable additivity, the condition will be satisfied.

In particular, if the sequence of the X_n converges (definitely) to a limit (in general random), $X = \lim X_n$, we can, if we assume countable additivity plus the uniform boundedness of the X_n, state that $\mathbf{P}(X) = \lim \mathbf{P}(X_n)$. The case of a series ΣX_h reduces to the preceding case if we consider the partial sums, $Y_n = \Sigma X_h$ ($h \leqslant n$). If we call Y' and Y'' the infimum and supremum of the sums, we have

$$\mathbf{P}(Y') \leqslant \inf \sum \mathbf{P}(X_h) \leqslant \sup \sum \mathbf{P}(X_h) \leqslant \mathbf{P}(Y''),$$

and, in particular, we have $\mathbf{P}(Y) = \mathbf{P}(\sum X_h) = \sum \mathbf{P}(X_h)$ if $Y' = Y'' = Y$ (i.e. the series is definitely convergent) under the condition that the remainders are uniformly bounded (and always, of course, with the assumption of countable additivity).[66]

We have said that we do not intend to consider countable additivity as an axiom; for us, it is a property that may appear more or less interesting and which will hold over certain partitions but not over others. Interpreting the condition as one of continuity, we can reformulate this fact in a more meaningful way by saying that it will hold *over certain linear spaces* and not over others.

This approach has the merit of spotlighting the real essence of the problem: the fact that the property of countable or finite additivity, that is of continuity or the absence of continuity, concerns the behaviour of the function **P** over a linear space L. To give a complete account of the behaviour of **P** in terms of continuity involves, therefore, distinguishing which linear spaces L belong to the complex Λ_p of linear spaces over which **P** is continuous, and which do not.

66 The convergence of the series $\sum \mathbf{P}(|X_h|)$ (together with the assumption of countable additivity) is a sufficient condition to establish that ΣX_h has probability = 1 of being convergent, and that (putting therefore, arbitrarily, $Y = Y'$, or $Y = Y''$, or $Y = Y' = Y''$, if they coincide, and otherwise $Y = 0$, etc.) one has $\hat{\mathbf{P}}(Y) = \sum \mathbf{P}(X_h)$ (N.B.: $\hat{\mathbf{P}}$ not **P**). Under this hypothesis, in fact, for arbitrary choices of positive ε and λ, there exists an N such that, for any q, we have $\sum \mathbf{P}(|X_h|)(N \leqslant h \leqslant N+q) < \lambda\varepsilon$, i.e. $\mathbf{P}\{\sum |X_h|(N \leqslant h \leqslant N+q)\} < \lambda\varepsilon$, and, *a fortiori* (denoting the preceding summation by $\{\Sigma_{\mathrm{N, q}}\}$ for short), $\mathbf{P}(\Sigma_{\mathrm{N, q}} > \lambda) < \varepsilon$ (because, if X is certainly positive, $(X > \lambda) \leqslant X/\lambda$). If the axiom of continuity holds, as we have assumed, the limit-event $(\Sigma_\mathrm{N} > \lambda) = \lim(\Sigma_{\mathrm{N, q}} > \lambda)$ (as $q \to \infty$) also has probability $\leqslant \varepsilon$, and, *a fortiori*, it follows that the fact that the series diverges (in which case the remainder Σ_N will be ∞) has probability $\leqslant \varepsilon$, and hence (since ε is arbitrary) *zero*.

More precisely, in line with what we have said previously, and with the condition of coherence, we shall say that **P** is *coherent and continuous* on L if no random quantity X of the form

$$X = k_1\big(X_1 - \mathbf{P}(X_1)\big) + k_2\big(X_2 - \mathbf{P}(X_2)\big) + \ldots + k_n\big(X_n - \mathbf{P}(X_n)\big) + \ldots$$

turns out to be uniformly positive (where the k_h are any real numbers and the X_h belong to $\mathscr{L}$), not only for sums involving a finite number of terms (as is required for coherence) but also for series (convergent, and with uniformly bounded remainders).

It is clear that if $\mathscr{L}$ belongs to Λ_p then so does every linear space contained in it, and so does the closure, $\bar{\mathscr{L}}$, formed by all the random quantities that can be obtained from $\mathscr{L}$ by means of the passage to the limit in the sense given:

$$\big(X_n \to X, |X_n - X| < K\big).$$

If $\mathscr{L}_1$ and $\mathscr{L}_2$ belong, then so does $\mathscr{L}_1 + \mathscr{L}_2$ (the linear space of sums $X_1 + X_2$, $X_1 \in \mathscr{L}_1$ and $X_2 \in \mathscr{L}_2$): this holds for any finite number of spaces L_h *but not for an infinite number.*[67] This indicates that the most 'natural' hypothesis is not true; a hypothesis which corresponds most closely to the standard point of view because it leads to a distinction between those events and random quantities which belong to a certain system of 'probabilizable' entities, and those which do not. This is the hypothesis that the complex Λ_p consists of all and only those linear spaces that belong to some given linear space L*, which, in this case, would have acquired the meaning of 'total field of continuity'.

19 Questions Concerning Qualitative Formulations

19.1. There are many senses in which the words qualitative probability have been used, some of them very different from each other. To attempt to list them and classify them would be both tedious and pointless, but something must be said in order to point out the necessity of not confusing things that do differ, and of not being put off by apparent absurdities. Among the latter, for example, we include the fact that one might expect to encounter rather vague considerations, but can, in fact, find oneself forced into hair-splitting detail, obliging one to apply, in all cases, the methods of comparison introduced for zero probabilities.

Our day-to-day judgements are on the whole rather vague and we usually limit ourselves to just a few verbal gradations (quite probable, or very, very much, not much, very little,…), or to percentage approximations (50%, 75%, 90%, 99%,…). In comparisons between two events, the probabilities will be said to be 'roughly equal' if the dominance of one over the other does not appear to be obvious. At this level, however, there is not even the possibility of arguing in mathematical terms.

67 Consider a countable partition of events E_{hk} ($h, k = 1, 2, \ldots, n, \ldots$). Let us denote by $E_h = \sum_k E_{hk}$ the sum of events whose first subscript is h, and suppose we attribute the values $p_{hk} = \mathbf{P}(E_{hk})$ and $p_h = \mathbf{P}(E_h)$ in such a way that $\sum_k p_{hk} = p_h$ (for each h), but $\sum_h p_h < 1$ (i.e. $\sum_{hk} p_{hk} < 1$). On the linear spaces $\mathscr{L}_h$ defined by the E_{hk} and E_h, **P** is continuous, and hence is also continuous on every linear space $\mathscr{L}$ determined by a finite number of L_h. This no longer holds, however, if we consider the space $\mathscr{L}$ determined by the whole infinite collection of L_h.

Sometimes, one thinks of vagueness in the sense of 'indeterminacy' (for example, between precise numerical bounds); we have already referred to this, and we shall return to it later. At other times, one is willing to compare (let us assume exact comparison, in order not to get lost in too many subcases) the probabilities of events but without using numerical probabilities. A physician might have a quite precise opinion, in a comparative sense, concerning the probabilities that a small number of patients will overcome their present disease, but without knowing what to do if he were required to compare them with the probability of obtaining something other than a '6' on the role of a die (or, more explicitly, were he required to state whether they were more or less than $\frac{5}{6}$; i.e. 83·3%). Sometimes, this inability to compare them with numbers is attributed to innate peculiarities of the events in question (rather than to contingent reasons, such as lack of practice; see Borel's review of Keynes' treatise[68]), or to the fact of not having at one's disposal (or not wishing to have) devices such as dice, urns and so on. In this case, if comparability is assumed exact, as in the comparison of intervals where one is led to say that a closed interval (i.e. end-points included) is greater than one of equal length but open (i.e. end-points excluded), it is clear that a non-Archimedean scale results (and this is the absurdity we referred to at the beginning).

Other considerations arise when indeterminacy has a precise meaning; when, on the basis of some data, one can establish only that a probability p belongs to an interval $p' \leqslant p \leqslant p''$. That we are not dealing with an essential indeterminacy is a point that we have stressed. Nevertheless, there is something to be said here and a few points will have to be made in connection with the discussion (in Sections 5–7) relating to the verifiability of events and measurement of quantities.

19.2. *Axiomatic formulations in qualitative form.* In all the methods of approach we have so far looked at, we have introduced, straightaway, numerical values for probabilities under the intuitive guise of prices, and as parameters required for optimal decision making. In so doing, we referred (albeit indirectly) to percentages of white balls, or to number of successes, and so on. This is certainly the most direct way of learning how to express one's own opinions and how to formulate the mathematical conditions which they must satisfy (and in terms of which they can be manipulated in a probabilistic argument).

There are occasions, on the other hand, when it seems preferable to start from a purely ordinal relation – that is a qualitative one – which either replaces the quantitative notion (should one consider it to be meaningless, or, anyway, if one simply wishes to avoid it), or is used as a first step towards its definition. For example, given two commodities (or two economic alternatives) A and B, one can ask which is preferable (or whether they are equally preferable) before defining utility (or perhaps even rejecting the very idea of measurable utility); and the same can be said for temperature, the pitch of a note, the length of intervals and so on.

One could proceed in a similar manner for probabilities, too. In fact (if one accepts the subjective point of view), one can apply precisely the same notion of preference as we mentioned for utility. Instead of two commodities A and B, one compares one and the same gain (let us say 1 lira) conditional on the occurrence of event A, or event B.

68 The article is reprinted in Borel's *Traité* (as note 2 in issue III of Vol. IV); an English translation is given in H.E. Kyburg and H.E. Smokler, *Studies in Subjective Probability*, John Wiley & Sons, Inc., New York (1964).

Our preference (apart from reservations concerning 'distorting factors'; see Section 13 above) will be for the event judged more probable (or, if the two events are judged to have the same probability, we will be indifferent).

This approach has been studied, and, provided one does not insist on splitting hairs, leads quickly and naturally to the usual conclusions (although in a form less directly applicable to the general case). The properties one needs to take as axioms are simple and intuitive (the standard order properties, plus the qualitative equivalent of additivity): given that E' and E'' are incompatible with E, then $E \vee E'$ is more or less probable than $E \vee E''$, or equally probable, according to whether E' is more or less probable than E'', or equally probable; in other words, logical sums preserve order.[69] All the same, the 'qualitative' comparison inevitably turns out to be far too precise (indeed, far too sophisticated), from a theoretical point of view, for what is required for the quantitative (numerical) evaluation; in any case, it is not conveniently translatable into such an evaluation (unless one considers the possibility of constructing special scales of comparison).

The complication derives from the fact that, in a qualitative sense, a possible event (no matter what probability p one attributes to it, even $p = 0$) is obviously 'more probable' than an impossible event. Similarly, by adding to an event E a possible event A, incompatible with it, even if of zero probability, one obtains an event $E + A$, which is 'more probable' than E. It follows that, having other events E' with (numerical) probabilities equal to that of E, $\mathbf{P}(E') = \mathbf{P}(E)$, the qualitative comparison would have to establish for each one whether it had the same probability as E, or $E + A$, or greater than the first and less than the second, or greater than both, or less than both. Even worse; consider an arbitrary sequence of events $A_1, A_2, \ldots, A_h, \ldots$, all of zero probability, mutually incompatible, and incompatible with E, and another sequence of events $B_1, B_2, \ldots, B_h, \ldots$, all of zero probability, mutually incompatible, and contained in E: setting

$$E_0 = E, \quad E_h = E + A_1 + A_2 + \ldots + A_h,$$
$$E_{-h} = E - B_1 - B_2 - \ldots - B_h, \quad (h > 0),$$

one obtains an increasing ($E_h \subset E_k$ for $h < k$) and doubly unbounded sequence of events E_h ($h = 0, \pm 1, \pm 2, \ldots, \pm n, \ldots$), all with probability $\mathbf{P}(E_h) = \mathbf{P}(E)$. Any comparison of an E' (also having probability $\mathbf{P}(E') = \mathbf{P}(E)$) with the E_h should make precise which (if any) of the E_h have the same probability as E'; or, otherwise, in which of the intervals E_h, E_{h+1}, it finds itself; or if it precedes, or follows, all the E_h, for h between $\pm\infty$.

The necessity, now explained, of this much more refined comparison, has led us to use the phrase 'having the *same* probability' for two events which, in the ordering, belong to the same 'equivalence class', rather than 'equally probable', which we use when we refer to the equality of their numerical probabilities.

The situation is that which would present itself (in a less serious way) in a comparison between intervals, if intervals of equal length were to be called 'equally long' only if they both contained either 0, 1 or 2 of their end-points (otherwise, the one containing more end-points would be called 'longer'). One arrives at a closer analogy by extending the example to sets which are unions of a finite number of intervals, and (the sum of the

---

69 See B. de Finetti, 'Sul significato soggettivo della probabilitá', *Fundamenta Mathematicae*, 17, Warsaw (1931); an improvement in the argument given in the notes of my course on the Calculus of Probability, University of Padua, 1937–1938, was made by Professor A. Gennaro who succeeded me in presenting the course.

lengths being equal) calling 'longer' the set for which the difference between the number of closed and open components is greatest (the intervals containing only one end-point are not counted; any isolated points are counted as closed intervals; these conventions are necessary if we are to have additivity, as in the case of probability).

Using such partitions into intervals as an image for our probabilistic partitions, one sees, for example, that, if the certain event is thought of as represented by a closed interval of length 1, it is impossible to divide it into two intervals (or, more generally, into n) that have the same probability (there are $n + 1$ end-points, one too many). This difficulty cannot be overcome by changing partitions into sums of intervals: it is always a question of dividing up the length 1 (into intervals with one end-point), plus one point. Shifting an end-point from one of the intervals to another one creates a disparity between them, but, in total, there always remains one end-point too many.

Conversely, if one does not consider that one has (included in the field of events to be compared) events that are suitable for furnishing a scale of comparison (for example, drawing balls numbered 1 to n, and judged to have the same probabilities, from an urn, with n arbitrarily large[70]), then the inequalities arising can provide totally inadequate information about the numerical values of the probabilities. Given, for example, a partition into three (incompatible) events A, B, C (assumed in order of decreasing probability), and that the only remaining comparison open to us is between A and $B + C$, this will tell us whether the probability of A is greater than or less than $\frac{1}{2}$. In the former case, we know only that $\mathbf{P}(A)$ lies between $\frac{1}{2}$ and 1, $\mathbf{P}(B)$ between 0 and $\frac{1}{2}$, $\mathbf{P}(C)$ between 0 and $\frac{1}{4}$; in the latter case, we know that $\mathbf{P}(A)$ lies between $\frac{1}{3}$ and $\frac{1}{2}$, $\mathbf{P}(B)$ between $\frac{1}{4}$ and $\frac{1}{2}$, $\mathbf{P}(C)$ between 0 and $\frac{1}{3}$ (Figure A.2). And it cannot be said that things necessarily improve if we consider more than three events. If, for example, the most probable of them is more probable than the union of the others, one can only say that its probability lies between $\frac{1}{2}$ and 1; the others, therefore, in aggregate, can have probability close to $\frac{1}{2}$, or arbitrarily close to zero, or even zero.[71]

We could avoid complications of this kind by assigning to the comparison 'A is more probable than B', a meaning equivalent to $\mathbf{P}(A) > \mathbf{P}(B)$, and, in particular, calling a possible event of zero probability 'equal in probability to the impossible event'. In order to do this, it would be necessary to introduce the Archimedean property; in other words, to characterize those events 'more probable than the impossible one' as those with a positive numerical probability by means of a condition like the following: 'there exists a finite N such that, in every partition into N events, at least one is less probable than the given event' (whose probability is then $\geqslant 1/N$). But why resort to this distortion of an arithmetic condition instead of proceeding directly, given that one's desired goal is, in fact, the arithmetic notion?

70 The inconvenience of having to postulate the existence of partitions into events having the same probabilities has been overcome in L.J. Savage, *The Foundations of Statistics* (Chapter 3, 3: 'Quantitative personal probability') by means of a weaker assumption: for any N, one can construct a partition into N parts such that no union of n parts is more probable than one of $n + 1$ parts (for any $n < N$).

71 In the case of $n > 3$ parts (and in the absence of a 'scale of comparison') it becomes complicated to even establish the compatibility of a system of inequalities (between sums of events of the partition). For $n = 4$, there is a simple sufficient condition (as I showed in my paper 'La logica del plausibile secondo la concezione di Pólya', *Atti. Riun. S.I.P.S.* 1949 (1951)). Contrary to what I had supposed, however, this does not hold for $n > 4$, as was shown by C.H. Kraft, J. Pratt and A. Seidenberg, 'Intuitive probabilities on finite sets', *Annals of Mathematical Statistics*, **XXX** (1959).

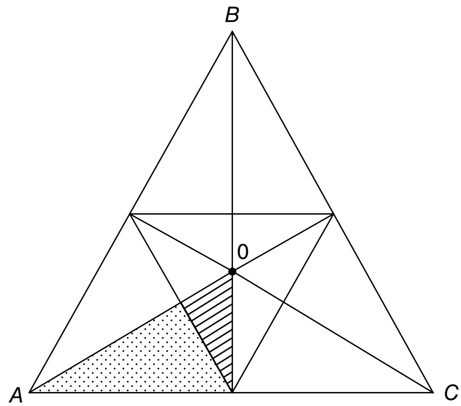

Figure A.2 Areas distinguished by means of the comparisons of probabilities for the events of a partition and their sums. The above refers to the case $n = 3$.

Dotted area: $A > B > C$, $A > B + C$.
Shaded area: $A > B > C$, $A < B + C$.

For $n = 4$ (tetrahedron), and $n > 4$ (simplex in higher-dimensional space), the mode of subdivision is similar.

On the other hand, the 'sophisticated', non-Archimedean, criterion corresponds exactly to the purely logical meaning that one would like to give to the comparison of probabilities in the particular case in which it relates to an objective condition, that of implication. If it is true (indeed, if it is certain; i.e. if we know) that $A \subset B$ (an objective condition), then coherence obliges us to evaluate $\mathbf{P}(A) \leqslant \mathbf{P}(B)$ (subjective evaluation). One can say that A is 'less probable in an objective sense' (or 'less possible') than B if and only if $A \subset B$, because if A occurs then B certainly occurs, and, moreover, it is possible that B occurs without A occurring. Viewed in this light, the fact that the event $B - A$ will be assigned zero probability under some evaluations, and nonzero probabilities under others, is irrelevant: the important thing is that $B - A$ is possible. To give a geometrical analogy in this case, also, one could say that a comparison between two sets in the absence of a notion of a metric can only lead one to assert that A is smaller than B when the former is properly contained within the latter. In this case, and only in this case, will it be true that $m(A) \leqslant m(B)$ for *whatever* measure m might be introduced (the exact form being $<$ or $=$, according to whether the measure in question attributes to $B - A$ a positive or zero value).

Given disparate events E' and E'', it seems too much to hope that by comparing their probabilities one can decide if $\mathbf{P}(E') = \mathbf{P}(E'')$ exactly, rather than whether there is a difference of about 10^{-6}, or 10^{-1000}, and so on. Should one wish to square the reasonableness of the procedure with the logical scruples mentioned above (i.e. the seemingly obvious fact that, for the same price, one prefers to have an extra possibility of winning, even though the extra possibility has probability zero), one could perhaps consider an intermediate order relation. One might define (for example):

$A < B$: 'A less probable than B'	if $\mathbf{P}(A) < \mathbf{P}(B)$, or if $\mathbf{P}(A) = \mathbf{P}(B)$ and $A \subset B$;
$A \sim B$: 'A not comparable with B'	if neither $A < B$ nor $B < A$; that is if $\mathbf{P}(A) = \mathbf{P}(B)$ and we have neither $A \subset B$ nor $B \subset A$.

Instead of following this rather abstract comparison of the various possibilities, it is more useful to ask oneself whether the exigencies of the problem themselves indicate the appropriate form, which could then profitably be adopted. This approach leads to something quite similar to the above, but, in the case where $\mathbf{P}(A) = \mathbf{P}(B)$, comparability will be given by a condition less restrictive than demanding that either $A \subset B$ or $B \subset A$ (i.e. either $A - AB = 0$, $\mathbf{P}(B - AB) = 0$, or vice versa). Specifically, the condition consists of $\mathbf{P}(A - AB) = \mathbf{P}(B - AB) = 0$, together with the comparability of these two zero probabilities, along the lines suggested in Chapter 4 and developed in Section 18.2 of this Appendix.

It is a question of comparing the conditional probabilities

$$\mathbf{P}\big(A - AB \,|\, A + B - AB\big) \quad \text{and} \quad \mathbf{P}\big(B - AB \,|\, A + B - AB\big)$$

(whose sum = 1), saying that A is more or less probable than B according to whether the first expression is greater than the second or is less. Note that in this way one can deal with every case within the one formulation; it does not matter whether $\mathbf{P}(A)$ and $\mathbf{P}(B)$ are different, or are equal, or whether, if they are zero, one has to proceed to the comparison of residuals. By definition, we have that A and B are not comparable if $\mathbf{P}(A) = \mathbf{P}(B)$ and the residuals $A - AB$ and $B - AB$ have equal probabilities (which is automatic if they are not zero).

Note also that it would be wrong to claim that events which are not comparable have the same probabilities. If A is more probable than B on account of the zero probability of $A - AB$ being greater than that of $B - AB$, and if C is an event such that $\mathbf{P}(C) = \mathbf{P}(A) = \mathbf{P}(B)$ but $\mathbf{P}(AC) < \mathbf{P}(C)$, then both A and B turn out not to be comparable with C; but to say 'having the same probability as C' would imply that they had the same probability as each other, and this is false by hypothesis.[72]

19.3. *Do 'imprecise probabilities' exist*? The question as it stands is rather ill-defined, and we must first of all make precise what we mean. In actual fact, there is no doubt that quantities can neither be measured, nor thought of as really defined, with the absolute precision demanded by mathematical abstraction (can we say whether the number in question is algebraic or transcendental? Or are we capable of giving millions of significant figures, or even a few dozen?). A subjective evaluation, like that involved in expressing a probability, attracts this criticism to an even greater degree (but this is no reason for regarding the problem differently in this case, as somehow being more essentially rooted in the concepts involved). The same is true for 'objective probability': the person putting his faith in 'objective probabilities' is in precisely the same situation, except insofar as he is restricting himself to cases in which everyone (he himself, or even a subjectivist) has at his disposal criteria and information which make the judgement easier.

In this sense, it should be sufficient to say that all probabilities, like all quantities, are in practice imprecise, and that in every problem involving probability one should provide, just as one does for other measurements, evaluations whose precision is adequate in relation to the importance of the consequences that may follow. In any case, one should take into account that there is always this margin of error (for instance, it might be worth repeating the calculations with several slightly different values).

---

72 The only case in which the two events E' and E'' could be said 'to have the same probability' is that in which they consisted, respectively, of m' out of n', and m'' out of n'', events of two partitions into events having the same probabilities, where $m'/n' = m''/n''$. This remark should not be taken to mean that we wish these futile and absurd complications to be taken seriously, but, on the contrary, that we wish to remove them, without ignoring, however, the issue of what can or cannot be expressed in a correct way.

The question posed originally, however, really concerns a different issue, one which has been raised by several authors (each of whom, it seems to me, imparts a different shade of meaning to the problem). It concerns the possibility of cases in which one is not able to speak of a single value p for a given probability, but rather of two values, p' and p'', which bound an area of indeterminacy, $p' \leqslant p \leqslant p''$, possessing some essential significance.

The idea can be traced back to Keynes (see the remark in the last section concerning Borel's review), and was later taken up by B.O. Koopman and I.J. Good, developed considerably by C.A.B. Smith,[73] and more recently by other authors, like Ellsberg and Dempster.

Several different situations may lead one to express oneself in terms of an imprecise evaluation.

An example of this occurs when one wishes to distinguish various hypotheses, and attributes different probabilities $\mathbf{P}(E|H_i)$ to an E, depending on the various hypotheses H_i; if one then ignores the hypotheses, one can only conclude that the probability lies between the maximum and the minimum. We have already dealt with this case in Chapter 4, Section 4.8, especially in 4.8.3 and 4.8.5. The probability is what it is on the basis of the information that one has. It is clear that with additional information the probability could take on all conceivable values, finally reaching, and then remaining at, either 1 or 0, when it is finally known whether it is true or false. If we are dealing with hypotheses H_i about which we expect soon to have some information, then it may be reasonable to wait until this information is available, rather than making a provisional evaluation by taking a weighted average with respect to the probabilities which, in the meantime, are attributed to the H_i. It would be naïve, however, to assert that $\mathbf{P}(E)$ will take on a value lying somewhere between the $\mathbf{P}(E|H_i)$. There is an infinite number of partitions in hypotheses and the information which comes along might be anything at all (for instance, it may confirm that out of the H_i, H'_j, H''_l those with $i = 3, j = 1, l = 7$ are true); the $\mathbf{P}(E \mid H_3 H'_1 H''_7)$ could vary anywhere between 0 and 1, even though the $\mathbf{P}(E|H_i)$ are all very close to one another (and even if they are equal; this case is only without interest insofar as the question would then, of course, never have been raised).

A second example occurs when one has not given sufficient thought to the matter, and hence possesses only a vague idea of the evaluation one wishes to make. A special case of this occurs when one has expressed the evaluation in terms of a formula (e.g. $p = \mathrm{e}^{-a}\mathrm{a}^{\mathrm{n}}/n!$, with $a = 5813$ and $n = 12$), but, having not yet carried out the numerical calculations, one has only a rough idea of the order of magnitude. In the final analysis, however, nothing has changed. One either carries out the calculations or one is obliged to take as the probability the prevision of the result according to one's own, more or less haphazard, crude estimation: there is no better solution.

The idea of translating the imprecision into bounds, $p' \leqslant p \leqslant p''$, even in the weaker sense proposed by Good (who regards p' and p'' not as absurd, rigid bounds, capable of

73 These questions are examined in detail in Sections 26 and 27 of the paper by B. de Finetti and L.J. Savage that we have frequently referred to; particular reference is made to the (then very recent) paper of C.A.B. Smith, and to the interesting discussion to which it gave rise at a Royal Statistical Society meeting, with contributions from Barnard, Cox, Lindley, Finney, Armitage, Pike, Kerridge, Bartlett and (in the form of a written contribution) Anscombe. The reader will find there many other points which we have not found room for here.

'making the imprecision precise', but merely as indications of maxima), is inadequate if one wishes to take it to the limit in the sense in which it serves to give an idea of the imprecision with which every quantity is known or can be considered. One should think of the imprecision in the choice of the function **P** (extended, for example, to some neighbourhood P* of a given $\mathbf{P}_0$ in the space P).The imprecision for individual events and random quantities would, as a consequence, be determined not as isolated features, but with the certain or uncertain connections deriving from logical or probabilistic relations.

19.4. *What one can do in practice.* A similar kind of discussion can be given concerning what one can actually do in practice; the main purpose of this, however, is to make clear that the issues involved here are rather different and do not give rise to any difficulties or anxieties.

The (theoretical) possibility of attributing a *precise probability* to all the events appears to be an indispensable requirement if one considers probability as a notion which applies to events per se, independently (see Section 3) of the existence and nature of any properties (e.g. topological) of the fields to which the definition of certain events can be referred. On the other hand, this does not imply that these probabilities are determined, as unique extensions of those conferring to a subfield (extensions which may or may not provide a sharpening of bounds; sometimes, as a special case, they may provide a unique value), nor does it imply that we are obliged to complete the evaluation, nor even to worry about it. Indeed, one can even call a halt well before this (earlier than usual) if there is no real interest in proceeding further: this is even more the case if the hypotheses upon which one would base oneself in proceeding further appear rather artificial and devoid of any realistic meaning.

If, in the context of the above (Section 19.3), we stop thinking in terms of a certain subset $\mathscr{P}^*$ of previsions $\mathbf{P} \in \mathscr{P}^*$, among which we are uncertain as to which one to choose, and we consider instead that we are dealing with the set of all the $\mathscr{P}$ which extend a given **P** in $\mathscr{L}$, then the bounds of indeterminacy mentioned above (at the end of Section 15) would follow. But it is not a question of imprecision. The fact that, in the case we are considering, it is only possible to say of a $\mathbf{P}(E)$ that it lies between $\mathbf{P}^-(E)$ and $\mathbf{P}^+(E)$, does not imply that certain events, like E, have an indeterminate probability: it merely implies that the probability is not uniquely defined by the initial data that one has considered. As an analogy, it would, in the same way, be nonsense to say that 'a rectangle having a perimeter of 12 m has an indeterminate diagonal'; it is 'determinate' in the sense that it is what it is, but one has to measure it, or measure a side, or something else, in order to obtain sufficient information to be able to 'determine' it (in the sense of obtaining, by means of a calculation, its well-determined value, notwithstanding the fact that knowledge of the perimeter alone is not sufficient).

There is a context in which one might refer to indeterminacy, but only in a precise, technical sense. This would arise if a certain individual were familiar with the evaluations that another individual had made within $\mathscr{L}$, and wished to establish what *the latter* should do outside that ambit in order to remain coherent. This makes it clear, however, that the indeterminacy is not a property of the events, but rather that it lies in the fact that an outsider cannot remove it, since he cannot replace the individual who is interested in the, as yet unprejudiced, evaluation.

Another aspect of the problem links up with the discussion (Sections 5–11) concerning the 'verifiability' of events (or the 'realizability' of measurements).

If, for a given problem, in a given situation, an event is not, in practice, verifiable, then any discussion about its probability is mere idle talk. For this reason, leaving aside the question of whether or not one accepts the necessity of admitting countable additivity as an axiom, it seems that if one is discussing the probability that $X \in I$, where I is non-measurable in the Jordan–Peano sense, and there is an interval in which both I and its complement $\tilde{I}$ are everywhere dense, then no measurement – not even with the assumption of 'unbounded precision' (Section 7) – can decide, on the basis of an observation x lying in that interval, whether or not the exact value lies in the interval. The same difficulty remains even under weaker assumptions: for example, if there exist points of both I and $\tilde{I}$ whose distance from x is smaller than the margin of measurement error.

In other fields, too, for realistic applications it seems much more useful to use methods which avoid too rigorous an assumption of precision. For example, when one speaks of 'convergence' it seems preferable, when considering results to have asymptotic validity, to confine attention to those which are valid for a large, but finite, number of cases (without taking seriously the notion of considering an infinite number of them).

20 Conclusions

Can one, having now come to the end, draw some conclusions?

I have in mind, of course, the critical questions that we have examined in this Appendix (some of which we anticipated in the text, to the extent that the topic in question required). In other words, I am referring to questions of a predominantly technical nature – if the word technical is adequate to characterize the difference between these matters and those concerning the meaning and formulation of the entire (subjective and Bayesian) theory; matters which occupied us throughout the text (Chapters 1–12).

Some kind of summary is required, if only to avoid the possibility of the reader being left with a feeling of confusion or bewilderment. The latter is a distinct possibility, because of the apparent contrast between our tendency on the one hand to simplify things, refusing to go beyond the level of practical applications, and towards, on the other hand, throwing ourselves headlong into hair-splitting and complicated analyses (which are not only far removed from any foreseeable application, but even strain the limits of good sense).

Why then, someone will surely ask, not be content with the 'happy medium' provided by the standard approach? This consists in proceeding to the point where countable additivity makes everything work beautifully, and then stopping when the miracle ceases.

Because – I would answer – so far as I am concerned it is by no means a 'happy medium,' but rather a case of 'two wrongs not making a right.' In my opinion, anything in the formulation that proceeds beyond what the Jordan–Peano–Riemann machinery provides is irrelevant for practical purposes, and unjustifiable on theoretical and conceptual grounds.

The two kinds of discussion to which we referred above, although apparently in contrast to one another, are intended, converging from opposite directions, to demonstrate one and the same point: *one can do without complications* (and this is perhaps the wisest course of action), but, should one decide to embark upon them, *one must do so wholeheartedly, in a constructive manner, even though this may prove troublesome.*

I may be wrong. My criticisms will not have been in vain, however, if in order to refute them someone comes forward and explains and justifies, in a sensible and meaningful way, those things which, up until now, have merely been 'Adhockeries for mathematical convenience.'