PATRICK SUPPES

THE ESSENTIAL BUT IMPLICIT ROLE OF MODAL CONCEPTS IN SCIENCE

When J. C. C. McKinsey and I were working on the foundations of mechanics many years ago, we thought it important to give a rigorous axiomatization within standard set theory, and we therefore resisted any use of modal concepts or counterfactual conditionals in the formulation of the axioms of mechanics. I continue to think that the use of an extensional set-theoretical framework is appropriate and adequate for most, if not all, scientific discourse. As my interests have shifted more to the foundations of probability and the applications of probability concepts in the behavioral sciences, however, I have gradually come to the position that modal concepts, especially as expressed in the use of probability concepts, are essential to standard scientific talk. Yet, in a majority of cases the modal concepts remain implicit in that talk, and their logic is scarcely used in either theoretical or experimental analyses of empirical phenomena. In this paper I expand on these main points under two headings, probability and physical space, each of which raises particular issues about modal concepts.

I. PROBABILITY

In discussing modal concepts in relation to probability concepts, I shall assume the standard background of set-theoretical probability theory as used, for example, in current work in mathematical statistics. I shall not talk about probability concepts as expressed by special languages of confirmation, which have received considerable attention from philosophers. The set-theoretical formulation I shall use is the familiar one of a probability space consisting of a sample space Ξ, a σ-algebra of subsets of Ξ and a probability measure on Ξ satisfying the usual measure-theoretic axioms.

To illustrate these concepts, it will be sufficient to consider a simple example. Let us look at the setup for flipping a coin three times. The sample space Ξ consists of eight experimental outcomes corresponding

Kenneth F. Schaffner and Robert S. Cohen (eds.), PSA 1972, 305–314.

to the eight possible sequences of heads and tails that may be observed in three trials. We can make a number of modal statements about this experiment in terms either of events or propositions. Because the language of events is more common than the language of propositions in standard probability theory, I shall use the former, but I see no conceptual obstacles to the direct translation from the language of events to the language of propositions. Under the ordinary and standard semantics we would give for this setup, the following statements would be true:

(1) It is possible that the event of getting heads on all three trials will occur.

(2) It is necessary that either at least one head or one tail occur in the three trials.

Other statements of possibility and necessity are easily constructed. On the other hand, such statements are ordinarily of little interest in a probability context, either for theoretical or for experimental purposes. The probability statements in which we are ordinarily interested, however, also have a clear modal character. For instance, let us assume that the probability measure on the sample space of our simple experiment is the standard one for Bernoulli trials with $p=\frac{1}{2}$. With the addition of this measure, we can make a number of additional statements of the following sort:

(3) The probability of getting a head on the second trial is independent of the outcome of the first trial.

(4) The probability of getting exactly two tails, given that at least two tails occur, is $\frac{3}{4}$.

It is clear how (4) and similar statements can be put explicitly into the language of events or the language of propositions:

(4′) The probability of the event of getting exactly two tails, given the occurrence of the event of at least two tails, is $\frac{3}{4}$.

(4″) The probability of the proposition that exactly two tails will occur, given that at least two tails have occurred, is $\frac{3}{4}$.

In remarking that the semantics of this setup is standard, I mean that the set of experimental outcomes corresponds to the set of possible worlds in the usual characterizations of the semantics of modal statements. Some qualification is perhaps needed on the use of the phrase 'usual characterizations', because the set of experimental outcomes in most cases of application is considerably more restricted than the set of possible worlds as ordinarily discussed in modal logic. For example, all set-theoretical relations are accepted as fixed in the set of experimental outcomes and are not subject to variation in possible worlds; whereas, in many cases in discussing the set of possible worlds, variations in everything but logical truths across possible worlds are permitted. Nevertheless, this possible difference does not affect seriously anything I have to say about the modal character of probability statements.

Another divergence from modal logic is apparent in the scientific practice of handling probability concepts. The probability measure and statements about probability are all handled in a purely extensional fashion, and the extensional status of probability concepts is no different from that of any other scientific concepts. In a natural way, a special modal status could be given to probability assertions, but it seems to me that a fair assessment of the practice indicates that the general tendency is exactly the opposite. This is one sense in which I claim that modal concepts are left implicit in science. I cite as a prime example the handling of probability assertions.

There is a deeper sense and perhaps a more important sense why modal concepts do not seem to have a prominent place in systematic formulation of scientific concepts and theories. This second sense derives from the fact that in most theoretical and applied uses of probability concepts of any complexity the set of experimental outcomes, that is, the sample space, is itself left implicit, and the analysis is wholly restricted to being formulated in terms of random variables. In standard mathematical language, a random variable is a real-valued function that is defined on a sample space and is measurable. The requirement of measurability is just that the set of points in the sample space for which the random variable is less than any given real number x is in the σ-algebra of the sample space. In other words, a random variable is defined with respect not only to a sample space Ξ, but also with respect to a given σ-algebra on Ξ.

This sounds as if we are simply extending the apparatus already introduced from sample spaces to functions defined on the sample space.

Much of the talk in standard probability texts reinforces this impression. However, a closer look at what is done and how random variables are used makes it clear that there is no real interest, either mathematically or empirically, in the character of the sample space. From a formal standpoint we obtain the distribution of a random variable in the standard fashion from the probability measure on the sample space on which the random variable is defined. But in practice we do not care about the sample space and deal directly with the random variable and its distribution. I emphasize that this is not simply a point of practice in applied use of probability, but it is also completely standard for mathematical work in probability theory as well.

A typical example of this language would be a standard formulation of the weak law of large numbers. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with mean m and finite variance. Then for any $\varepsilon > 0$

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \leqslant \varepsilon\right) = 1.$$

In similar fashion stochastic processes are defined not in terms of sample spaces, but in terms of sequences of random variables (in the discrete case) and as an indexed family of random variables in the more general case. The set of possible experimental outcomes is not of interest and is not considered explicitly, or, to put it in modal terms, the set of possible worlds is not of interest and is not explicitly introduced in the apparatus of analysis. A somewhat amusing point on these matters is that, contrary to the kind of notation logicians would most like ordinarily, the very notation used for talking about the probability of a random variable's exceeding a certain value, etc., or of having a value lying in a certain interval, does not itself use a notation for the arguments of the random variable, but only for the random variable itself. Thus, while the logical tendency is to define the notation $P(X \geqslant x)$ by $P\{\xi : X(\xi) \geqslant x\}$, the notation without the explicit use of the arguments of the function is the standard practice.

What I am saying then is that in practice the modal character of probability statements, as well as reference to the set of possible experimental outcomes or worlds, is omitted, and the modal character of probability assertions is left implicit.

This elimination of the sample space, and thereby the set of possible worlds, is not a casual accident, but rather it reflects a natural scientific tendency to eliminate excess baggage. In practice, the sample space itself plays no direct role, and consequently, there is no need to use it explicitly in the formulation of probability statements. One of my main theses about the status of modality in scientific contexts is that there is this strong tendency to leave modal concepts implicit, because their explicit use does not seem to have an effective role to play in formulating theory or reporting experimental results.

One line of talk in statistics that might seem contrary to the use of probability concepts I have described is the sometimes confusing discussions of finite samples and the infinite populations from which they are claimed to be drawn. In applied statistics books there is indeed, I think, confusion on these matters and a tendency to feel that a badly explained modal concept of 'infinite population' has been introduced. I believe, however, that this can be straightened out in relatively simple terms by using the concept of random variable, distinguishing the sample distribution of a random variable from the theoretical or population distribution, and talk about drawing a sample from an infinite population of men, balls, urns or the like is not needed. In any case, the confusion that sometimes obtains in the talk about *sample* and *population* does not obviate my main point that modal concepts are left implicit.

It is worth examining further what a systematic approach to random variables would look like that avoids the underlying sample space of possible worlds. A first thought might be just to identify the random variable with its distribution. But it is clear that this proposal is not satisfactory, because any two random variables with the same distribution would be identical, and the completely natural idea of a random variable's representing a distributed quantitative attribute or property of a set or 'population' of objects would be lost.

A tentative proposal that would not require any substantial change in the current mathematical formulation of theorems about random variables is the following.

A *random framework* is an ordered quintuple (Ξ, U, X, F^*, F), where: Ξ is a finite set, the actual population of objects or events being studied; U is a sampling distribution on Ξ, in simple cases the uniform distribution; X, the random variable of the framework, is a real-valued function defined on Ξ; F^* is the frequency distribution of X, necessarily a step function because of the finiteness of Ξ; and F is the theoretical distribution of X that can in principle be constructed by smoothing from F^* but that in practice has a fixed number of parameters that can be estimated from the sample distribution of a sample drawn from Ξ according to the sampling distribution U.

As in current practice the mathematical study of random variables would concentrate just on the pair (X, F). Probability statements about the random variable X are defined in terms of F, and the finiteness of the domain of X is ignored. Thus,

$$P(X \leqslant x) = F(x),$$
$$P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1),$$

and so forth. Since F is a distribution function, it is a nondecreasing function of x, continuous to the right and such that $F(-\infty)=0$ and $F(\infty)=1$, and consequently, such probabilities as $P(X<x)$ are also well defined in terms of F.

There is, of course, from an intensional standpoint a residual objection to this proposed definition of random frameworks and random variables. It is that two distinct properties, e.g., height and weight, could in principle lead to extensionally identical frameworks. A way of reducing if not entirely eliminating this problem is to require that the random variables take physical quantities rather than real numbers as their values. (A detailed and explicit mathematical theory of physical quantities is to be found in Chapter 10 of Krantz *et al.*, 1971.)

In the present context, however, a more interesting proposal is to define random variables as quantitative properties together with their distributions and to consider only the pair (X, F). Two random variables (X_1, F_1) and (X_2, F_2) are then identical under this characterization if and only if X_1 and X_2 are the same properties, i.e., are intensionally identical, and F_1 and F_2 are extensionally identical as distribution functions. In this setup, probability statements about X_1 or X_2 are defined as above in terms of F, and, as before, X_1 and X_2 continue to be real-valued functions defined for a given set of objects or phenomena.

It may be claimed that with this proposal the set of possible worlds is re-introduced in terms of the semantics of property terms. Certainly this is a defensible viewpoint and I can only say that the role of possible worlds is now more implicit than in the case of the sample-space formulation of random variables. Unlike the sample-space approach, the set of possible worlds no longer enters directly into the relevant mathematical definitions of probability statements about random variables. The definitions given above in terms of the distribution function provide a self-sufficient mathematical basis. I may seem to be unduly belaboring the status of random variables as objects, but their widespread use in scientific practice, coupled with a history of confused identity would seem to warrant this detail.

II. SPACE

Much of what I have said about probability concepts also holds, in my judgment, for spatial concepts. The intellectual framework for talking about space, however, is rather different from that for talking about probability, and consequently, what I want to claim about spatial concepts needs its own set of conceptual defenses. I shall conclude by saying that there is a much closer affinity than many people would be prepared to admit between the probabilistic concept of sample space and the concept of physical space.

When we examine the usual axioms for Euclidean geometry there seems to be nothing of a modal character about them. If we take elementary axioms that depend upon the primitive concepts of betweenness and the quaternary relation of equidistance, i.e., $xyEuv$ iff the distance between x and y is equal to the distance between u and v, then the axioms fall into two parts. In the first part we have universal axioms that apply to any points without restriction, and in the second part we have existential axioms asserting that if certain points are given, then certain other points must exist.

The explicit formal statement of axioms for elementary geometry, as in Tarski (1959), requires no modal concepts in the language of the theory. In fact, the proof of various metamathematical facts about elementary geometry, for example, the existence of a decision procedure, depends upon staying within a well-defined framework of first-order extensional logic with identity.

On the other hand, from a more general standpoint, we can regard the axioms of geometry as having as their domain possible objects rather than real objects. We can challenge the real existence of the points postulated, and thereby of the other geometric figures that are brought into existence, so to speak, by the existence of the standard manifold of points.

It seems philosophically straightforward, even if ultimately it is not, to trêat this vast manifold of points in the same fashion as the set of all denumerable sequences of ones and zeros, as in the sample space for the experiment of flipping a coin an infinite number of times. Roughly speaking, we construct the sample space by listing the two possible outcomes for a given trial and then taking the denumerable Cartesian product of that set of outcomes. What is important in comparison with the axioms for Euclidean geometry is the absence in the description of this sample space of any existential statements.

The second important point to note is that while the sample space of denumerable sequences is, of course, never realized in an actual experiment, it is an idealization convenient for mathematical purposes in analyzing limiting properties of the sort already exemplified in the weak law of large numbers stated above. Such remarks about an infinite sample space are obvious and familiar. It is less obvious and less familiar to try to make similar claims about physical space. It is these claims that I want to expand upon here.

The first question to consider is how to eliminate the existential axioms characteristic of standard formulations of Euclidean geometry and how to use axioms that are in the spirit of the sample-space construction just characterized. A natural approach, I believe, is to base the axioms of geometry on quantifier-free operations. In Moler and Suppes (1968), axioms for constructive geometry are given in terms of the operation of finding the intersection of the two line segments determined by four points, and the operation of laying one line segment off on another. In order not to have one-point models of the axioms, it is necessary to assume also as part of the primitive concepts three distinct points α, β and γ that are noncollinear and that do not form an equilateral triangle. (This is the assumption required for two-dimensional geometry; four points that are noncoplanar and that do not form a regular tetrahedron are required in the three-dimensional case.) The three distinguished points, α, β and γ, play the role in the sample-space construction of the set of outcomes – heads and tails – of a given trial. The reiterated application of the two basic operations of intersection and laying-off of segments leads to the construction of new points and corresponds approximately to the operation of forming the denumerable Cartesian product in the sample-space construction.

Let me say at once that I do not think the constructive geometry formulated by Nancy Moler and me is really in the best form to exhibit what I call here the 'sample-space' nature of physical space. There are several reasons for this. One natural restriction would be to build in more explicitly a theory of actual bodies and their motions. The purely geometric operations we considered do not provide a satisfactory means for representing bodies and their motions. Second, by picking two line segments that are almost parallel, we get at once the existence of points at an indefinitely large distance from our place of initial operation. It would be a reasonable, but not necessary, condition on the construction of physical space as the set of possible points or possible paths to require that the operations unfold from the region of initial given points in a manner that seems more finite and regular.

I shall not attempt to examine the technical details of constructing the space-time manifold as the set of possible paths of particles or bodies. The simpler constructive geometry, based upon the two operations of intersection and laying-off of segments, provides a conceptual model of the kind of construction I have in mind. As we consider the closure of these two operations under denumerable iteration, we obtain a Euclidean-like space of infinite extent that consists of a denumerable number of points. (From a mathematical standpoint, the space is isomorphic to a vector space over a Pythagorean field, but this point is not essential here.) Just as in the case of the sample space for infinite sequences of coin flips, we can regard this constructive space of points as the sample space of possible points realizable by constructive operations. And, just as in the case of the sample-space construction, we can ask ourselves whether it is reasonable to drop the set of possible points and be concerned only with objects that correspond to the random variables discussed earlier. I do not think the analogy is perfect, but I want to examine it, because to the extent that it holds, it reinforces my thesis that we have a natural tendency to give modal concepts an implicit rather than explicit role in scientific theory construction.

In Euclid, and more generally in ancient Greek geometry, there is no mention of the space of points, and we can do physics in the same spirit.

As Herman Weyl (1922, p. 152) succinctly put the matter many years ago,

> Only the motions of bodies (point-masses) relative to one another have an objective meaning.

We do not and we cannot directly observe pure space-time points, or even invariant relations holding between them simply as space-time points.

The thrust of my remarks is to make bodies play the role in physics that random variables play in probability theory, and to make the underlying space implicit in both cases. What I have not tried to deal with here is how to interpret the global properties of space that arise in general relativity theory. I have not thought the matter through in a satisfactory fashion, but I would like to conclude by conjecturing that the kind of viewpoint I am advocating can be developed in a self-consistent fashion. The manifold of space-time points is regarded as the manifold of possible paths of bodies, and this manifold is itself defined as the asymptotic limit of certain constructive operations performable on bodies.

Stanford University

BIBLIOGRAPHY

Krantz, D. H., Luce, R D., Suppes, P., and Tversky, A., *Foundations of Measurement*, Vol. 1. Academic Press, New York, 1971.

Moler, N. and Suppes, P., 'Quantifier-Free Axioms for Constructive Plane Geometry', *Compositio Mathematica* **20** (1968) 143–152.

Tarski, A., 'What is Elementary Geometry?', in L Henkin, P. Suppes and A. Tarski (eds.), *The Axiomatic Method*, North-Holland, Amsterdam 1959.

Weyl, H., *Space-Time-Matter*, 4th ed., Methuen, London 1922.