Foundations of Physics, Vol 30, No 10, 2000

Invariance, Symmetry and Meaning[1]

Patrick Suppes[2]

Received October 18, 1999

The role of the concept of invariance in physics and geometry is analyzed, with attention to the closely connected concepts of symmetry and objective meaning The question of why the fundamental equations of physical theories are not invariant, but only covariant, is examined in some detail The last part of the paper focuses on the surprising example of entropy as a complete invariant in ergodic theory for any two ergodic processes that are isomorphic in the measure-theoretic sense.

1. COMMON MEANING OF INVARIANCE

The ordinary or common meaning of invariance gives a very nice sense, qualitatively, of its technical meaning. Something is invariant if it is not varying, unalterable, unchanging, or constant. Of course, the first question that then arises, is this: "What is unalterable or unchanging?" The intuitive answer, which again matches well, in a general way, the technical answers I shall examine later, is that a property of an object, collection of objects or more generally, some phenomena, is invariant. Familiar examples are the shape of a cup as we move the cup around, rotate it, turn it upside down, etc. The shape of the cup does not change, nor does its weight. An important physical example is that the speed of light in a vacuum is invariant with respect to any inertial frame of reference with which we measure it, contrary to the simple addition of velocities familiar in classical physics

It is the idea of invariance expressing constancy that is especially important. Thus, we say, psychologically, someone's attitudes are invariant

[1] It is a pleasure to dedicate this article to Maria Louisa Dalla Chiara, from whom I have learned a lot about the foundations of physics in lectures and conversations over many years

[2] C S L I , Ventura Hall, Stanford University, Stanford. California 94305-4115

0015-9018/00/1000-1569$18 00/0

or unchanging, just because they are constant over the years. Of course, all of the meanings I have stated are approximate synonyms for each other Something that is constant is unalterable Something that is constant is unchanging and so forth.

2. GENERAL LOGICAL RESULT ON INVARIANCE

A classical result expresses very well the important idea that that which is invariant with respect to any one–one mapping of the world, or universe of any model under consideration, is just the logical relations. Let me give two formulations, one that goes back to the early work of Lindenbaum and Tarski (1932–1933/1983, p. 385): "Roughly speaking, Th. 1 states that every relation between objects (individuals, classes, relations, etc.) which can be expressed by purely logical means is invariant with respect to one–one mapping of the 'world' (i e., the class of all individuals) onto itself and this invariance is logically provable. The theorem is certainly plausible and had already been used as a premiss in certain intuitive considerations. Nevertheless it had never before been precisely formulated and exactly proved."

A closely related but different formulation is given much later by Tarski and Givant (1985, p. 57): "(i) Given a basic universe U, a member M of any derivative universe $U^{\sim}$, [e.g., the Cartesian product $U \times U$] is said to be logical, or a logical object, if it is invariant under every permutation P of U On the basis of (i) one can show, for example, that for every (nonempty) U there are only four logical binary relations between elements of U: the universal relation $U \times U$, the empty relation $\varnothing$, the identity relation $U \mid Id$, and the diversity relation $(U \times U) \cap Di$."

What this latter formulation brings out is that there are only four logical binary relations. No other binary relation is invariant, that is, constant, under arbitrary permutations of the universe. Later I shall consider the geometric program of invariance with which the nineteenth-century mathematician Felix Klein is closely associated. The Tarski results quoted represent the most general expression of that approach to invariance.

3. SYMMETRY

Closely associated with invariance is symmetry, as has been stressed by many commentators on invariance The example closest to us is the approximate bilateral symmetry of the human body. It is sometimes plausibly argued, although I shall not do so here, that the psychological

appeal of bilateral symmetry in designs, and in architectural structures, is a consequence of our own bilateral symmetry. Whether this psychological claim is true or not is not easy to determine, but the appeal of bilateral symmetry in structures and in design is very evident. I show just two figures here. One is a Chinese design (Jones, 1867). The vertical line bisecting the picture is the axis of bilateral symmetry. The second figure is of a plan and elevation of a villa by Palladio (Palladio, 1570/1731). The plan and elevations drawings, even more than the structure itself, make evident the effort to have as much bilateral symmetry as possible.

We can generalize from bilateral symmetry to the geometric viewpoint derived from Felix Klein. It can be stated this way. "For a given group of transformations of a space, find its invariant properties, quantities or relations." Tarski's theorem gives this result for the most general transformations, in the sense of all one–one transformations on a space. More particularly, for a given geometry, which, from the standpoint formulated

Fig 1 A Chinese design

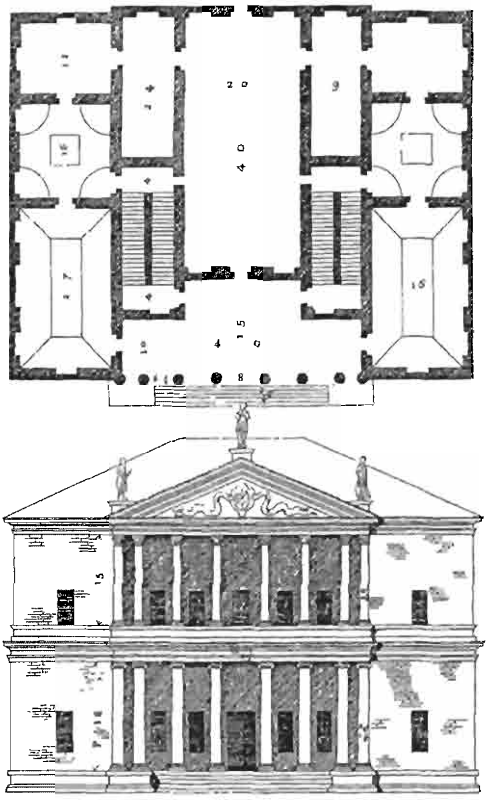

Fig 2 Plan and elevation of a villa by Palladio

here, means a set-theoretical structure consisting of the basic set and the relations and functions on this set, a transformation that carries one geometric structure into another, on the same space, is an *automorphism* of the space and Klein's viewpoint was that the group of automorphisms of a geometry defines the geometry Or, put in a more purely Kleinian way, any group of automorphisms of a space defines a geometry and there will be relations and functions of this geometry that are invariant under the group of automorphisms.

Within this framework, we say that a geometric figure is *symmetric* under a given group if every automorhphism of the group maps the figure into itself. For example, a bilaterally symmetric figure is so mapped by a mirror reflection along what we familiarly call a line of symmetry, in the case of plane figures. Figures symmetric under various rotations are also familiar examples of symmetry, the circle above all, but, of course, the

square is also symmetric under any 90° rotation. Notice the difference. The square has a very small finite group of rotations under which it is invariant, whereas the circle is invariant under an infinite group of rotations.

In similar fashion three-dimensional figures like spheres and cubes are also invariant under various groups of motions or transformations. Idealization is important here. Most real physical objects look, or are different, in some observable, even if minute, way, under rotations or reflections.

Other familiar examples of geometric invariants are the relation of betweenness for affine geometry. Taken together, betweenness and congruence are complete invariants for Euclidean geometry, and therefore form a set of concepts in terms of which Euclidean geometry can be axiomatized.

It is worth remarking why symmetry and groups naturally go together. We may think of it this way. A geometric figure F in a space A is *symmetric* under a single transformation φ mapping A onto A if φ maps F onto itself. Now if this holds, so should the inverse mapping φ^{-1}, since $\varphi^{-1} \circ \varphi$ is just the identity map of A. In addition, it is natural to expect closure of this symmetry property. If φ_1 and φ_2 both map F onto itself then so should their composition $\varphi_1 \circ \varphi_2$. These two properties of a set G of transformations:

(i) If $\varphi \in G$ then $\varphi^{-1} \in G$,

(ii) If $\varphi_1, \varphi_2 \in G$ then $\varphi_1 \circ \varphi_2 \in G$,

are sufficient to guarantee G is a group of transformations or automorphisms of the given space A, since composition of transformations is necessarily associative, and $\varphi \circ \varphi^{-1}$ is the identity. This same argument also shows why invariance and groups so naturally go together.

When artists or architects draw a symmetric figure or design, it is undoubtedly usually the case that they do not explicitly think in terms of the automorphisms of the space in which the figure or design is implicitly embedded, and which map the figure onto itself. Even sophisticated mathematicians of ancient Greek times did not have an explicit group concept, but once uncovered, it seems a natural formalization of the very intuitive idea of symmetry.

4. MEANING

It is easy to pass from symmetry to meaning. Here are some familiar examples Orientation, such as Aristotle's up and down for space, is not meaningful in Euclidean geometry. Why? It is not invariant under the Euclidean group of motions, i.e., rotations, translations and reflections. In

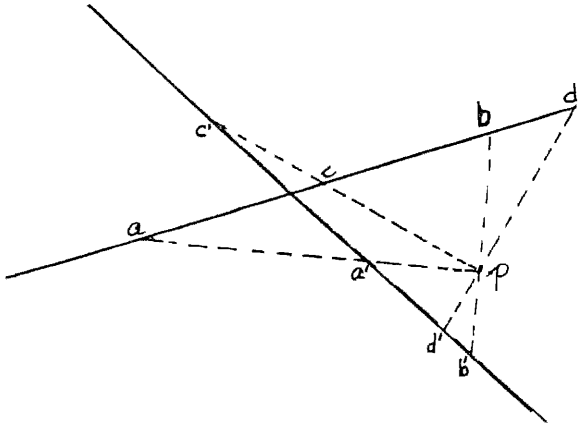

Fig 3 Invariance of separation under projection

general, a distinction that is not meaningful in a given geometry, or, more generally, with respect to a group of transformations, is the source of a symmetry.

In Euclidean geometry it is meaningful to compare the lengths of two line segments, regardless of whether they are parallel or not. In affine geometry for such a comparison to be meaningful, however, the line segments must be parallel, which includes lying on the same line. It is a distinguishing feature of affine geometry, in fact, that it is not a metric space There is no common measure along lines as we change direction from one line to another. It is only parallel segments that have a meaningful concept of congruence, which is what we mean when we speak of affine congruence.

In projective geometry, even the affine concept of betweenness is not preserved under projections from one line to another from a given point of perspective. In this case, we must go to a four-place relation. The relation of separation: $ab\ S\ cd$ if and only if a and b separate c and d, and conversely. Figure 3 shows how separation is preserved under projection, but betweenness is not. For example, we can easily see in the figure that b lies between a and d, but the projection b' does not lie between a' and d'. On the other hand, the pair ab separates the pair cd, and conversely; moreover, under projection the pair $a'b'$ separates the pair $c'd'$, and conversely.

5. OBJECTIVE MEANING

In physics, there is a tendency to go beyond simply talking about meaningfulness and to talk about *objective* meaning. Here the important

concept is that we are interested in invariant relations, or invariance more generally, for observers moving relative to each other with constant velocity. So, for example, in classical physics, two inertial observers should confirm the same measurements of that which is genuinely invariant. This doesn't mean that they don't make correct measurements in their own frames of reference, using their individual measurement procedures. But, as we know, directions, for example, will be different for the two observers, as will other properties. On the other hand, what we expect to hold is that the distance between two simultaneous spatial points is the same for all observers using the same calibration of their measurement instruments. The same is true of time. Temporal intervals are the same for all inertial classical observers using clocks with the same calibration. As is emphasized in foundational discussions of classical physics, spatial and temporal intervals are observer-independent, but the concepts of the center of the universe, the beginning of time or being in a state of absolute rest are not invariant and thus have no physical meaning, in spite of earlier views, such as those of Aristotelian physics, to the contrary. Of course, the use of these concepts by Aristotle was in no sense ridiculous. Phenomenological experience certainly gives us a natural concept of absolute rest, i.e., of being at rest with respect to the earth, and the same for the center of the universe. On the other hand, as a matter much farther from familiar experience, Aristotle famously argued that the world is eternal, i.e., time has no beginning, creating problems for later Christian theologians.

Here is a second example from special relativity. Invariance for two inertial observers of the "proper" time of a moving particle is sufficient to derive that their observations are related by a Lorentz transformation. The proper time interval τ_{12} of two points on the trajectory of an inertial particle satisfies the following equation:

$$\tau_{12}^2 = (t_1 - t_2)^2 - \frac{1}{c^2}\left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]$$

The important concept here is that proper time has objective meaning independent of the observer. Moreover, proper time has the important feature of being a complete invariant, since its invariance determines the group of Lorentz transformations (Suppes, 1959)

A third example is the classic theorem of Noether (1918). To every one-parameter group of diffeomorphisms of the configuration manifold of a Lagrangian mechanical system which preserves the Lagrangian function, there corresponds an invariant quantity, which is a first integral of the equations of motion.

Example A. Translation of the system in a given direction preserves the Lagrangian function. Then the center of mass moves in this direction with constant, i.e., invariant, velocity.

Example B. A system admits rotations around a given line. Then the angular momentum with respect to this line as axis is invariant, i.e., constant, in time.

These various geometric and physical examples make clear that the concept of invariance is relative not only to some group of transformations, but more fundamentally and essentially, relative to some theory. Invariant properties of models of a theory are the focus of investigations of invariance. So, for example, Aristotle's theory of the heavens led naturally to their being eternal, since they were unchanging, and consequently there could be no beginning of time.

An invariance holding across many theories is the implicitly or explicitly assumed flatness of space in Ptolemaic, Copernican, Galilean, Newtonian and Einsteinian special-relativity theories of the physical universe. A technical geometric way of formulating the matter is the fundamental assumption that space-time is a four-dimensional affine space, which has many invariant consequences.

6. WHY THE FUNDAMENTAL EQUATIONS OF PHYSICAL THEORIES ARE NOT INVARIANT

Given the importance, even the deep physical significance, attached to the concept of invariance in physics, and often also in mathematics, it is natural to ask why are theories seldom written in an invariant form

My answer will concentrate on physics, and somewhat more generally, on theories of measurement. One obvious and practically important point in physics is that it is simpler and more convenient to make and record measurements relative to the fixed framework of the laboratory, rather than record them in a classical or Lorentzian invariant fashion.

Moreover, this point is even more obvious if we examine the use of units of measurements. Analysis of measurements of length and mass will suffice In both these cases, selection of the unit of measurement is arbitrary, i.e., not physically invariant. What is invariant is the ratio of distances or masses The awkward locutions of continual use of ratios rather than the shorthand of conventional standards of units can be seen in ancient Greek mathematics. In fact, elementary computations show the advantage of units. Suppose we need to compare about 1000 distances

between cities with airports. If only invariant ratios were recorded, then about half a million ratios as pure numbers would be required, as opposed to 1000 physical quantities, i.e, 1000 numbers with a common unit, such as meters or kilometers.

The passion for writing ratios in the Greek style lasted a long time. Newton's *Principia* (1686/1946, p. 13) is a fine example. Here is Newton's classic geometrical formulation of his second law of motion. "The change of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed." It is often claimed, and quite sensibly, that moving away from the Greek geometrical style of expression to algebraic and differential equations was of great importance in introducing more understandable and efficient methods of formulation and computation in physics.

Apart from the rather special case of ratios, the same distinction is evident in synthetic and analytic formulations of standard geometries. A synthetic and invariant formulation of Euclidean geometry is easily formulated, just in terms of the qualitative relations of betweenness and congruence The analytic representations, in contrast, are relative to an arbitrary choice of coordinate system and are therefore in and of themselves not invariant. An invariance theorem is ordinarily proved to show how the various appropriate analytic representations are related, i.e., by which transformations. These transformations, in fact, constitute the analytic form of the group of Euclidean motions already mentioned.

The advantages of the analytic representations, particularly in terms of vector spaces, is universally recognized in physics It would be considered nothing less than highly idiosyncratic and eccentric for any physicist to recommend a return to the synthetic geometric formations used before the eighteenth century

Put in this transparent geometric context, it can be seen immediately that there is an inevitable conflict between invariance, as in synthetic formulations, and efficient computations, as in analytic formulations. In fact, selection of the "right" coordinate system to facilitate, and particularly simplify, computation is recognized as something of significance to be taught and learned in physics.

Beyond Symmetry. From what I said it might be misleadingly inferred that in choosing convenient laboratory coordinate systems physicists neglect considerations of invariance. In practice, physicists hold on to invariance by introducing and using the concept of covariants. Before introducing this concept, it will be useful to survey the natural generalizations beyond symmetry in familiar geometries.

The most general automorphisms of Euclidean space are transformations that are compositions of rotations, reflections and translations. We

have already considered rotations and reflections. Familiar figures such as circles, squares and equilateral triangles are invariant under appropriate groups of rotations and reflections—an infinite group for circles, and finite groups for squares and equilateral triangles.

On the other hand, none of these familiar figures are invariant under automorphisms that include translations. In fact, no point is, so that the only invariant is that of the entire space being carried onto itself. But translations are important not only in pure geometry but in many applications, especially in physics, where the Galilean transformations of classical physics and the Lorentz transformations of special relativity play an important role, and both include translations

In the last paragraph an unnoted shift was made from automorphisms of a qualitative form of Euclidean geometry, based on such primitive concepts as betweenness and congruence, to the numerical coordinate frames of reference of the Galilean and Lorentz groups of transformations. Such frames are needed to give an explicit account of the physicists' concept of covariants of a theory.

Typical examples of such covariants are velocity and acceleration, both of which are not invariant from one coordinate frame to another under either Galilean or Lorentzian transformations, because, among other things, the direction of the velocity or acceleration vector of a particle will in general change from one frame to another. (The scalar magnitude of acceleration is invariant.)

The laws of physics are written in terms of such covariants Without aiming at the most general formulation, the fundamental idea is conveyed by the following Let $Q_1, \ldots, Q_n$ be quantities that are functions of the spacetime coordinates, with some Q_i's being derivatives of others, for example. Then in general, as we go from one coordinate system to another, $Q'_1, \ldots, Q'_n$ will be covariant, rather than invariant, and so their mathematical form is different in the new coordinate system. But any physical law involving them, say,

$$F(Q_1, \ldots, Q_n) = 0 \tag{1}$$

must have the same form

$$F(Q'_1, \ldots, Q'_n) = 0 \tag{2}$$

in the new coordinate frame. This requirement of same form is the important invariant requirement. Equations (1) and (2) are, in the usual language of physicists also called covariant. So, the term *covariant* applies both to quantities and equations. I omit an explicit formal statement of these familiar ideas.

Here is a simple example from classical mechanics. Consider the conservation of momentum of two particles before and after a collision, with v_i the velocity before, w_i the velocity afterward, and m_i the mass, $i = 1, 2$, of each particle. The law, in the form of (1), looks like this:

$$v_1 m_1 + v_2 m_2 - (w_1 m_1 + w_2 m_2) = 0$$

and the transformed form will be, of course,

$$v_1' m_1 + v_2' m_2 - (w_1' m_1 + w_2' m_2) = 0$$

but the forms of v_i and w_i will be, in general, covariant rather than invariant. The masses m_1 and m_2 are, of course, invariant.

Finally, I mention that in complicated problems, proving that a physical law is covariant, i.e., the invariance of its form, can be, at the very least, quite tedious.

7. ENTROPY AS A COMPLETE INVARIANT IN ERGODIC THEORY

I now turn to an extended example that is one of the most beautiful cases of invariance, with consequences both in mathematics and in physics. We will need to build up some apparatus for this discussion.

Let us first begin with a standard probability space $(\Omega, \mathfrak{I}, P)$, where it is understood that $\mathfrak{I}$ is a σ-algebra of subsets of Ω and P is a σ-additive probability measure on $\mathfrak{I}$. We now consider a mapping T from Ω to Ω. We say that T is *measurable* if and only if whenever $A \in \mathfrak{I}$ then $T^{-1}A = \{\omega : T\omega \in A\} \in \mathfrak{I}$, and even more important, T is *measure preserving* if and only if $P(T^{-1}A) = P(A)$. T is *invertible* if the following three conditions hold: (i) T is $1-1$, (ii) $T\Omega = \Omega$, and (iii) if $A \in \mathfrak{I}$ then $TA = \{T\omega : \omega \in A\} \in \mathfrak{I}$. In the applications we are interested in, each ω in Ω is a doubly infinite sequence and T is the *right-shift* such that if for all n, $\omega_n = \omega'_{n+1}$ then $T(\omega) = \omega'$. Intuitively this property corresponds to stationarity of the process—a time shift does not affect the probability laws of the process, and we can then use T to describe orbits or sample paths in Ω.

I now introduce the central concept of entropy for ergodic theory. To keep the mathematical concepts and notations simple, I shall restrict myself to discrete-time processes and, in fact, to processes that have a finite number of states, as well. So, for such processes we have a simple definition of entropy.

First, the entropy of a random variable **X** with a (discrete) probability density $p(x)$ is defined by

$$H(\mathbf{X}) = -\Sigma p_i \log p_i$$

In a similar vein, for a stochastic process $\chi = \{\mathbf{X}_n : 0 < n < \infty\}$

$$H(\chi) = \lim \frac{1}{n} H(\mathbf{X}_1, \ldots, \mathbf{X}_n)$$

Notice that what we have done is define the entropy of the process as the limit of the entropy of the joint distributions. To illustrate ideas here, without lingering too long, we can write down the explicit expression for the joint entropy $H(\mathbf{X}, \mathbf{Y})$ of a pair of discrete random variables with a joint discrete density distribution $p(\mathbf{XY})$.

$$H(\mathbf{X}, \mathbf{Y}) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

which can also be expressed as

$$H(\mathbf{X}, \mathbf{Y}) = -E \log p(\mathbf{X}, \mathbf{Y})$$

where E is the expectation in this case, of the given function of the random variables **X** and **Y**. Notice, of course, there is a requirement that I have been a little bit casual about. It is important to require that the limit exist. So the entropy of the stochastic process will not be defined if the limit of the entropies of the finite joint distributions does not exist as n goes to infinity

For a Bernoulli process, that is, a process that has on each trial identically and independently distributed random variables (i.i.d.), we have a particularly simple expression for the entropy. It is just the entropy of a single one of the random variables, as the following equations show:

$$H(\chi) = \lim_{n \to \infty} H(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$$

$$= \frac{nH(\mathbf{X}_1)}{n} = H(\mathbf{X}_1) = -\sum p_i \log p_i$$

We require only a slightly more complex definition for the case of a Markov chain, again, of course, with the provision that the limit exists

$$\begin{aligned} H(\chi) &= \lim H(\mathbf{X}_n \mid \mathbf{X}_{n-1}, \ldots, \mathbf{X}_1) \\ &= H(\mathbf{X}_2 \mid \mathbf{X}_1) \\ &= -\sum_i p_i \sum_j p_{ij} \log p_{ij} \end{aligned}$$

Now we give, without detailed technical explanations, because the examples will be simple, the definition of ergodicity. Let $\chi(\Omega, \mathcal{F}, P, T)$ be a probability space with measure-preserving transformation T. Then T is *stationary* if for any measurable set A, $\mu(TA) = \mu(A)$, and transformation T is called *ergodic* if for every set A such that $TA = A$, the measure of A is either 0 or 1. We obtain a stochastic process by defining the random variable $\mathbf{X}_n$ in terms of T^n, namely, for every ω in Ω, $\mathbf{X}_n(\omega) = \mathbf{X}(T^n\omega)$. It is easy to see that Bernoulli processes as defined above are stationary and also ergodic. (A more intuitive characterization of stationarity of a stochastic process is that every joint distribution of a finite subset of random variables of the process is invariant under time translation. So, for every k,

$$p(x_{t_1+k}, x_{t_2+k}, \ldots, x_{t_m+k}) = p(x_{t_1}, x_{t_2}, \ldots x_{t_m})$$

Perhaps the simplest example of a nonergodic process is the Markov chain in heads and tails:

	h	t
h	1	0
t	0	1

Given that the initial probability $p_1(h) = p_1(t) = \frac{1}{2}$, the event A consisting of the sample path having heads on every trial has probability $\frac{1}{2}$, i.e., $p(A) = \frac{1}{2}$, but $T(A) = A$ and thus the process is not ergodic. It is intuitively clear, already from this example, what we expect from ergodic processes: a mixing that eliminates all dependence on the initial state.

On the other hand, we have an ergodic process, but the entropy is 0, when the matrix is slightly different and the probability of heads on the first trial is now 1.

	h	t	
h	0	1	$p(h_1) = 1$
t	1	0	

It is clear enough from the one possible sequence *hththt* . that all measurements or outcomes are predictable and so the entropy must be 0 Note what this process is. It is not a Bernoulli process, but a very special deterministic Markov chain with two states. In connection with this example, I mention that the standard definition in the literature for ergodic Markov chains does not quite require stationarity. The definition ordinarily used is that a Markov chain is ergodic if it has a unique asymptotic probability distribution of states independent of the initial distribution. So it is the independence of initial distribution rather than stationarity that is key. Of course, what is obvious here is that the process must be asymptotically stationary to be ergodic.

There are some further important distinctions we can make when a stochastic process has positive entropy; not all measurements are predictable, but some may be We can determine, for example, a factor of the process, which is a restriction of the process to a subalgebra of events. A factor is, in the concepts used here, always just such a subalgebra of events. The concept of a K process (K for Kolmogorov), is that of a stochastic process which has no factor with 0 entropy, that is, has no factor that is deterministic. Bernoulli processes are K processes, but here are many others as well.

One important fact about ergodic processes is the fundamental ergodic theorem. To formulate the theorem we need the concept of the indicator function of a set A. Let for any event A, I_A be the indicator function of A That is, for $\omega \in A$

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

The meaning of the fundamental ergodic theorem is that the time averages equal the space or ensemble averages, that is, instead of examining a cross-section of time for all ω's having a property expressed by an event A, we get exactly the same thing by considering the shifts or transformations of a single sample path ω. The theorem holds for all ω's except for sets of measure 0

Theorem. For almost any ω in Ω

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k \omega) = P(A)$$

Isomorphism of Ergodic Processes We define two stationary stochastic processes as being *isomorphic (in the measure-theoretic sense)* if we can map

events of one to events of the other, while preserving probability. The intuitive meaning of this is that their structural uncertainty is the same.

More technically, we define isomorphism in the following way. Given two probability spaces $(\Omega, \mathfrak{I}, P)$ and $(\Omega', \mathfrak{I}', P')$ and two measure-preserving transformations T and T' on Ω and Ω' respectively, then we say that $(\Omega, \mathfrak{I}, P, T)$ is *isomorphic in the measure-theoretic sense* to $(\Omega', \mathfrak{I}', P', T')$ if and only if there exists a function $\varphi\colon \Omega_0 \to \Omega'_0$, where $\Omega_0 \in \mathfrak{I}$, $\Omega'_0 \in \mathfrak{I}'$, $P(\Omega_0) = P(\Omega'_0) = 1$, that satisfies the following conditions:

(i) φ is $1-1$,

(ii) If $A \subset \Omega_0$ and $A' = \varphi A$ then $A \in \mathfrak{I}$ iff $A' \in \mathfrak{I}'$ and if $A \in \mathfrak{I}$

$$P(A) = P'(A')$$

(iii) $T\Omega_0 \subseteq \Omega_0$ and $T'\Omega'_0 \subseteq \Omega'_0$,

(iv) For any ω in Ω_0

$$\varphi(T\omega) = T'\varphi(\omega)$$

The mapping of any event A to an event A' of the same probability, i.e., condition (ii), is conceptually crucial, as is the commutativity expressed in (iv).

A problem that has been of great importance has been to understand how stochastic processes are related in terms of their isomorphism. As late as the middle fifties, it was an open problem whether two Bernoulli processes, one with two states, each with a probability of a half, $B(\frac{1}{2}, \frac{1}{2})$, and one with three states, each having a probability of one third, $B(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, are isomorphic, in accordance with the definition given above. The problem was easy to state, but difficult to solve. Finally, in the late fifties, the following theorem was proved by Kolmogorov and Sinai.

Theorem 1 (Kolmogorov, 1958, 1959; Sinai, 1959) If two Bernoulli processes are isomorphic then their entropies are the same.

Since $B(\frac{1}{2}, \frac{1}{2})$ and $B(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ have different entropies, by contraposition of Theorem 1, they are not isomorphic. It was not until twelve years later that Ornstein proved the converse.

Theorem 2 (Ornstein, 1970). If two Bernoulli processes have the same entropy they are isomorphic.

With these two theorems, extension to other processes, in particular, Markov processes, was relatively straightforward. And so, the following theorem was easily proved.

Theorem 3. Any two irreducible, stationary, finite-state discrete Markov processes are isomorphic if and only if they have the same periodicity and the same entropy.

Out of this came the surprising and important invariant result. Entropy is a complete invariant for the measure-theoretic isomorphism of aperiodic, ergodic Markov processes. Being a complete invariant also means that two Markov processes with the same periodicity are isomorphic if and only if their entropy is the same. What is surprising and important about this result is that the probabilistic structure of a Markov process is large and complex. The entropy, on the other hand, is a single real number. It is hard to think of an invariance result for significant mathematical structures in which the disparity between the complexity of the structures and the easy statement of the complete invariant is so large. Certainly, within probability theory, it is hard to match as an important invariance result. It is also hard to think of one in physics of comparable simplicity in terms of the nature of the complete invariant.

On the other hand, it is important to note that even though a first-order ergodic Markov process and a Bernoulli process may have the same entropy rate and, therefore, be isomorphic in the measure-theoretic sense, they are in no sense the same sorts of processes. We can, for example, show by a very direct statistical test whether a given sample path of any length, which is meant to approximate an infinite sequence, comes from the Bernoulli process or the first-order Markov process. There is, for example, a simple chi-square test for distinguishing between the two. It is a test for first-order versus zero-order dependency in the process.

But this situation is not at all unusual in the theory of invariance. We can certainly distinguish, as inertial observers, between a particle that is at rest in our inertial frame of reference, and one that is moving with a constant positive speed. What is observed can be phenomenologically very different for two different inertial observers, Galilean or Lorentzian. This way of formulating matters may seem different from the approach used in the geometrical or ergodic cases, but this is so only in a superficial way. At a deeper and more general level, all the cases of invariance surveyed here have the common theme of identifying that which is constant or unchanging.

REFERENCES

O Jones, *The Grammar of Chinese Ornaments* (Gilbert, London, 1867)

A N Kolmogorov, "A new metric invariant of transient dynamical systems and automorphisms in Lebesgue space," *Dokl Akad Nauk SSSR* **119**, 861 (1958), (Russian) MR 21 #2035a

A N Kolmogorov, "Entropy per unit time as a metric invariant of automorphism," *Dokl Akad. Nauk SSSR* **124**, 754 (1959), (Russian) MR 2 #2035b

A Lindenbaum and A Tarski, *Über die Beschränktheit der Ausdrucksmittel deductiver Theorien* (Ergebnisse eines mathematischen Kolloquium, fascicule 7 (1934/35)), pp 15–22, translated by J H Woodger, in *Logic, Semantics, Metamathematics*, 2nd edn, J Corcoran, ed (Hacket, Indianapolis, 1983), pp 384–392

I Newton, *Principia* (1686); translated by F Cajori (University of California Press, Berkeley, 1946)

E Noether, "Invarianten beliebiger Differentialausdrucke," *Nachrichten von der Gesellschaft der Wissenschaften zu Gottingen* (1918), pp 37–44

D S Ornstein, "Bernoulli shifts with the same entropy are isomorphic," *Adv. Math* **4**, 337 (1970)

A Palladio, *The Four Books of Architecture* (1570), translated by Isaac Ware (London, 1731)

Y G Sinai, "On the notion of entropy of a dynamical system," *Dokl Akad Nauk SSSR* **124**, 768 (1959)

P Suppes, "Axioms for relativistic kinematics with or without parity," in *The Axiomatic Method with Special Reference to Geometry and Physics*, L Henkin, P Suppes, and A Tarski, eds (Proceedings of an international symposium held at the University of California, Berkeley, December 16, 1957–January 4, 1958) (North-Holland, Amsterdam, 1959), pp. 291–307

A Tarski and S Givant, *A Formalization of Set Theory without Variables* (American Mathematical Society, Providence, Rhode Island, 1985)