SPACE, TIME, and SPACETIME

Lawrence Sklar

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY • LOS ANGELES • LONDON

TO LISSY

"Did I ever tell you my theory of Time and Space?"
They ignored him.

John Gardner,
The Wreckage of Agathon.

University of California Press
Berkeley and Los Angeles, California
University of California Press, Ltd.
London, England

First Paperback Edition, 1977
ISBN: 0-520-03174-1
Library of Congress Catalog Card Number: 73-76096
Printed in the United States of America

2 3 4 5 6 7 8 9 0

spacetime of the world is like and, once again on general relativity's account, just why it is the way it is. Much that I say will have to be taken on faith for the time being. But the reader can have some hopes that what he is forced to accept on authority here will be given argumentative support in Chapter IV.

1. *Minkowski Spacetime*

We can begin the construction of the Minkowski spacetime, the spacetime appropriate to the special theory of relativity, and the comparison of it with the prerelativistic notions by making some remarks about the "basic individuals" of which "space" (in the abstract sense) is constructed.

I have been talking previously of a space as being a set of *points*. But for spaces that are supposed to be more than mathematical abstractions, which are actually, somehow or other, to represent the world, what should we choose as the *points?* There are importantly-controversial philosophical issues involved here; I shall be discussing them at length toward the end of Chapter II and throughout Chapter III. For the moment, though, I shall try to ignore these issues. I will frame what I have to say in a language that, although it *suggests* a definite philosophical viewpoint about the reality and nature of the points, is not meant here to have any deep philosophical import or to beg any important questions.

It is convenient, even in characterizing prerelativistic theories, to start off with a relativistic technique. Let us imagine some idealized *events* in the world, for example, the collision of two perfectly unextended point masses or the intersection of two nonparallel perfectly breadthless light rays. We will take an idealized event as "marking" a definite location in spacetime. It will be convenient, however, to have such locations where events do not ever occur; so we will use a trick, at least as old as Leibniz, and speak not only about actual idealized events but about possible ones as well. The points of spacetime, then, will be all the locations of possible ideal events. Since the events are extensionless, so are their locations. We will adopt a traditional, but misleading, language and speak of the locations of possible idealized events as themselves being *events*. Spacetime is, then, the set of all events. But this is not meant to imply that real events, or even possible events, are the individuals of the set which is space. Rather, these individuals are *locations* of possible events which we are now calling 'events.'

The fundamental assumption of prerelativistic spacetime theory is that these fundamental individuals, the events, are not truly fundamental. Every event location can, in fact, be "analyzed" as an ordered pair of individuals, a spatial location and a temporal moment. Prerelativistically, what is the location of an idealized event? It is a specification of two "independent" features: (a) where the event took place, i.e., the spatial location of the event, and (b) when the event took place, i.e., at what moment of time it occurred. So, if e is an event location, we can look upon e as $e = \langle p,t \rangle$ where p is a place and t is a moment of time. Given the fundamental notions of places and times, instead of the set of events we can consider the set of places and the set of temporal moments. We can describe the structure of these sets and we can then *reconstruct* the structure of the spacetime set by examining its "composition" out of space and time.

According to the Newtonian theory, the structure of P, the set of spatial locations, is simple. It is E^3, i.e., Euclidean three-space. The structure of the set of temporal moments, T, is even simpler; it is E^1, the structure of the one-dimensional real line. The individuals of spacetime, events, are ordered pairs of places and times. So the structure of the spacetime is simply $E^3 \times E^1$, the Cartesian product of space and time, or the set of ordered pairs, the first element of which is selected from E^3 and the second from E^1.

We can extract the spatial structure from the spacetime structure as follows: Take spacetime as a class of events and as having the structure $E^3 \times E^1$. Define a relation among events, 'sim,' where e_1 sim e_2 if and only if the temporal location of e_1 equals the temporal location of e_2. 'Sim' will be an *equivalence* relation, that is, a transitive, reflective, symmetric relation on the set of events. We can use it to split up the set of events into "equivalence classes," where two events are in the same class if and only if they occur at the same time. The equivalence classes generated out of the spacetime by 'sim' will all have the structure E^3. Intuitively, 'sim' is the relation that holds between two events when they are *simultaneous.* An exactly parallel procedure, using the equivalence relation 'occur at the same place' instead of 'occur at the same time,' will split up the spacetime into a collection of sets of events at a single location, all of which have the structure E^1.

Given a pair of events, e_1 and e_2, what significant questions can we ask about them in this picture of spacetime? (A) We can ask if they occurred at the same time. We can ask, given that they did not occur at the same time, just how large the temporal interval between them was. (If they did not occur at the same time we can also ask which of them was *later* than the other. I will forgo probing into this question until Chapter V.) (B) We can ask whether or not the events occurred at the same place. If they did not, we can inquire about the spatial separation between them. Reflection on the spacetime structure we have been examining will show that the question, "How far apart in space were e_1 and e_2?" is a meaningful question whether or not e_1 and e_2 were *simultaneous* events. We shall later see, throughout Chapter III, how this feature of what I will call from now on 'Newtonian spacetime' led to controversy during the prerelativistic period of science, and how it still has some puzzlement attached to it.

We can now construct *Minkowski spacetime* by contrasting it with the Newtonian spacetime I have just described. Once again, the spacetime is a set of events. But the structure imposed upon this set is radically different from the one I have just described. First, some remarks about what Minkowski spacetime is *not:* (1) It does not have the structure $E^3 \times E^1$. (2) It is not meaningful to ask for the temporal separation between two events, or even to ask whether or not the events are simultaneous. (3) It is not meaningful to ask for the spatial separation between two events, or even if the two events occurred at the same place.

What, then, is the structure of Minkowski spacetime and what questions have meaning relative to this structure?

Minkowski spacetime is a space (a) that has four dimensions; (b) that is a differential manifold and an affine space; (c) that has the *topological* structure of E^4, the Euclidean four-dimensional space; but (d) that has a radically different metric structure from E^4. In Minkowski spacetime we do not discuss distances between events, but rather the *interval* between them. The interval between two events along a curve, A, is defined like this:

1. $\text{interval}\,(P,Q) \text{ along } A = \int_A ds$
2. $ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 - c^2\, dt^2$

where c is the velocity of light in a vacuum and is a constant. If we take $c = 1$, which is simply a matter of adopting new units for space (or time) measurement, equation (2) simplifies to $ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 - dt^2$.

Let me expand upon the meaning of equation (2). Once again, let me warn the reader. At this point I am simply trying to briefly describe the spacetimes appropriate for relativistic theories. He shall have to forgo a rationalization for adopting such a spacetime model until Chapter IV.

Minkowski spacetime is a set of events. It is four-dimensional. Between any two events of the spacetime, and along any curve, there is a quantity, the interval between the events along the curve, which is a number and which is an *invariant* property of the spacetime. The spacetime is such that between any two events there is a curve such that the interval between these events along this curve has an *extreme* value, relative to the interval between the events along any neighboring curve. That is, $\text{int}^2(P,Q)$ along this curve is either (*a*) greater than $\text{int}^2(P,Q)$ along neighboring curves, or (*b*) less than $\text{int}^2(P,Q)$ along neighboring curves, or (*c*) $\text{int}^2(P,Q)$ along the curve is zero while it is positive or negative along all neighboring curves. These "extremal curves" are the geodesics of Minkowski spacetime.

The interval between events along a curve is not a *distance* between them, for distances are always nonnegative whereas the square of the interval can be positive, zero, or negative. If the square of the extremal interval between P and Q is positive, we say that P and Q have *spacelike separation,* if it is negative we say that they have *timelike separation,* and if it is zero we say that P and Q have *lightlike separation* (Fig. 18). The set of events that have lightlike separation from a given event, e, is called the *light-cone for* e. Any set of events which is a submanifold of the space and all of whose point events have spacelike separation from e is called a *spacelike hypersurface for* e.

We can characterize some of the physical meaning of these relations. If P and Q have lightlike separation, then a light ray emitted from one of the two events can reach the other, or else P and Q are the same event (remember that P and Q are events, *not spatial locations*). If P and Q have timelike separation, then a causal signal whose velocity of propagation is less than that of light can get from one of these events to the other. If the events have spacelike separation, then no causal signal can get from one event to the other unless its velocity of propagation exceeds that of light in a vacuum. Since, as we shall see in detail in Chapter IV, a fundamental assumption of relativistic theories is that there are no such faster-than-light signals, spacelike-separated events have no possible causal signal connecting them at all.

The extremal interval between events is an invariant. The spatial

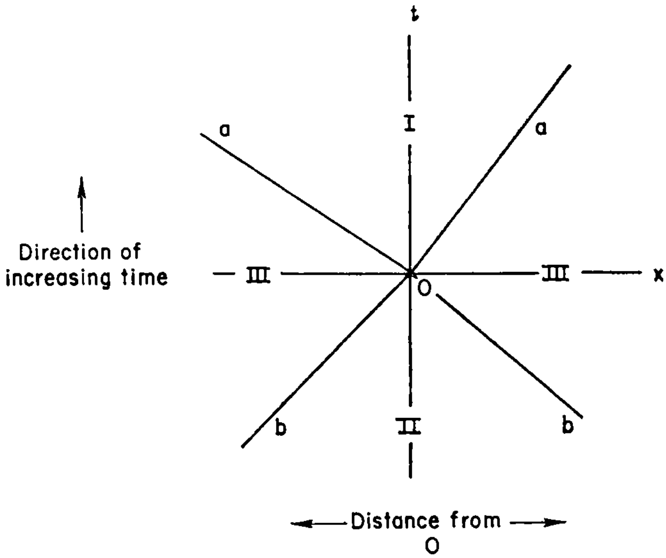

Fig. 18. Some very basic elements of Minkowski spacetime. In this diagram, two spatial dimensions have been suppressed, so properly speaking it is a diagram of a Minkowski spacetime with only one spatial dimension.

The lines a represent light signals sent from O, and the events along these lines are those with lightlike separation from O in the future time direction. The events along the lines b are those with lightlike separation from O in the past time direction.

The events in region I are those with timelike separation from O in O's "absolute future," and those in region II are the events with timelike separation from O in O's "absolute past."

The events in regions III are those with spacelike separation from O, i.e., those not connectible to event O by any causal signal whatever.

separation and the temporal separation are *not*. We can get a notion of spatial and temporal separation, however, by introducing a coordinate system. Two basic features of Minkowski spacetime relevant to coordinatizing are these: First, since the spacetime is four-dimensional, we can supply an internal coordinatization that labels each point, i.e., each event, by four numbers. Call these x^1,x^2,x^3 and $t = x^4$. Second, although quite arbitrary coordinatizations are possible, the spacetime is such that certain particularly simple coordinatizations can be found. These "special" coordinatizations are closely analogous to the Cartesian coordinates imposable on Euclidean n-space by the means of n mutually perpendicular straight lines. The fact that such a coordinatization can be found for Minkowski spacetime indicates something important about it, its "flatness," just as the fact that Cartesian coordinates can be imposed on a Riemannian n-dimensional space indicates that the space is a Euclidean flat n-dimensional space.

As we shall see in Chapter IV in detail, relativistic spacetimes are characterized by the fact that their splitting up into space and time is dependent upon the state of motion of the "observer." That is, two observers in motion with respect to one another will, in general, disagree about the spatial and temporal separation of a pair of events but will agree about the interval between the events. We shall also see that for special relativity there are observers whose states of motion are distinguished from those of observers in general. These are the observers who are in uniform or "inertial" motion. It is these special observers who will be able to naturally impose upon the events of the spacetime their "preferential" coordinatization. This is the coordinatization in which the "differential of the interval" has the form given it in equation (2) above. Relative to it, the equation of the geodesic connecting spacelike-separated events is a linear equation, just as the equation for a straight line is a linear equation in a Euclidean n-space when we impose upon the space a set of arbitrary Cartesian coordinates. For an inertial observer, the spacetime does split up into $E^3 \times E^1$, but this splitting is by no means invariant—which events go into which spaces with which other events varies from observer to observer. But, relative to a given observer, the splitting up can be done. The result is a relativized spacetime that can, like the nonrelativized spacetime of Newtonian theory, be viewed as "Euclidean three-space persisting through time." The spatial three-spaces can be Cartesian coordinatized, as can the temporal one-space, and the result is a coordinate system in which $ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 - (dx^4)^2$.

The close analogy between Euclidean four-space and Minkowski spacetime leads us to call Minkowski spacetime a *pseudo-Euclidean* space. *Euclidean* because Cartesian coordinates can be imposed upon it once an observer in a particular inertial state of motion is selected; *pseudo* because the invariant quantity relating pairs of events is not the always nonnegative distance invariant of Euclidean n-space, but instead, the invariant interval of Minkowski spacetime whose square can be positive, zero, or negative, depending upon whether the events are spacelike-, lightlike-, or timelike-separated from one another.