

The Causal Theory of Space-time

“It is not sufficient to say that Einstein’s clocks and measuring rods are *ideal* ones: for, before we are in a position to speak of them as being ideal, it is necessary to have some clear conception as to how one could, at least theoretically, recognize ideal clocks or measuring rods in case one were ever sufficiently fortunate as to come across such things; and in case we have this clear conception, it is quite unnecessary, in our theoretical investigations, to introduce clocks or measuring rods at all.”

A. A. Robb

I. Introduction

The special theory of relativity forced a radical revision of classical views about the causal structure of the world. One of the most obvious changes involved the possible rates of causal propagation. According to the special theory there is a finite upper limit to the speed of causal chains, whereas classical causality allowed arbitrarily fast signals. Foundational studies, such as that of Reichenbach (1969), soon revealed that this departure from classical causality in the special theory is intimately related to its most dramatic consequences: the relativity of simultaneity, time dilation, and length contraction. By now it has become clear that these kinematical effects are best seen as consequences of the geometrical structure of Minkowski space-time, which in turn incorporates a nonclas-

NOTE: I am greatly indebted to Adolf Grünbaum, Allen Janis, David Malament, Howard Stein, and John Stachel for many helpful remarks and discussions. Geoffrey Matthews (Indiana University) is responsible for many of the details of the construction of section III. Above all, I wish to thank my colleague at Indiana University, J. Alberto Coffa, for many invaluable discussions, decisive criticisms, and an extremely careful and critical reading. As a result of his efforts, this essay certainly contains less error, and, one hopes, more of the truth. To be sure, any errors that remain are entirely my own responsibility (unless they were overlooked by Coffa). Finally, I would like to thank Loretta Moses for her patience and skill in typing this manuscript.

This research was supported, in part, by National Science Foundation grant SOC73-05753.

sical theory of causal structure. However, it has not been widely recognized that the converse of this proposition is also true: *the causal structure of Minkowski space-time contains within itself the entire geometry (topological and metrical structure) of Minkowski space-time.*

Most likely, this remarkable result has been overlooked for a variety of reasons. First of all, it is by no means obviously correct; its demonstration requires a good deal of logical and mathematical effort, and until recently, the existing proofs—in the works of Robb (1914, 1936)—have lacked economy and elegance. Second, the causal definability of the geometry of flat space-time is in direct conflict with the view that the determination of geometry *necessarily* involves the introduction of metrical conventions. Finally, the inability of *classical* causality to provide more than merely an account of temporal *order* provides plausible grounds for regarding this as a fundamental limitation of *any* causal theory; thus while special relativity has been cited by defenders of a causal theory of time as providing fresh support for the causal theory of *temporal order*, these same causal theorists have not seen that the causal resources of this theory are so powerful as to yield in addition a purely causal account of temporal metric, spatial geometry, and the topology and metric geometry of Minkowski space-time. Even in general relativity, causal structure determines the topological, differentiable, and conformal structures of a wide class of physically “reasonable” space-times; the metric of these space-times, however, is not causally determinate (see section VIII below).

In the following section, the classical causal theory is outlined and its ability to provide a basis for the construction of time and space is investigated. It is shown that the classical theory of causal structure permits at most the construction of an (absolute) temporal *order*; the source of this limitation of classical causality is exhibited. Section III introduces the causality-related notions to be used throughout, developing these from the single symmetrical relation of causal connectibility. In section IV, the causal theory of the *topology* of Minkowski space-time is then outlined, and its adequacy as a basis for the topology of the manifolds of general relativity is discussed briefly. In section V the linear structure of Minkowski space-time is constructed out of causality; Robblike methods of developing its metrical structure are sampled. These constructions illustrate the way in which the causal theory of Minkowski space-time bridges the gap between “qualitative” and metrical properties of space-time. Adequacy proofs for the causal construction of the linear structure of Minkow-

ski space-time are provided. In section VI it is shown how the causal cone, together with the affine structure of Minkowski space-time, may be used to define its standard metrical structure: congruence, orthogonality, and the space-time interval. In section VII, the causal definability of congruence in Minkowski space is again demonstrated using group-theoretical methods and a theorem of E. C. Zeeman (1964). The construction of the linear structure of Minkowski space on this basis is more or less immediate, and is indicated in outline. Finally, in section VIII, some objections to the causal theory are considered, and the extent to which the causal theory may be extended to general relativity is discussed.

II. Classical Causality and Its Limits

Leibniz seems to have been the first to conjecture that the structure of time and space might be reducible to causality (1956), and the basic ideas of his causal theory are both simple and ingenious. The *simultaneity* of two distant events is defined as their lack of causal interaction. Thus if we choose a given event and consider all those events in the universe neither affecting nor affected by the chosen event, we hope to obtain an *acausal* "slice" of the world: a set of events no two of which causally interact. Identify such a slice with an instant of time, and define time itself as the set of all such slices or instants. Now order the slices or instants using the causality relation between events chosen from within each slice. Instant *A* (slice *A*) is then prior in time to instant *B* (slice *B*) just in case every event in *A* causally influences every event in *B*. In this way the temporal order of the world is reduced to causal relations between its events.

So far, only the order *between* acausal slices has been mentioned. What of the order *within* each slice? This, according to Leibniz, is where space appears in the construction. The events within each acausal slice have an order, presumably the same for all slices, and this order is just the *spatial* order of the world.

Even this rough sketch of the causal theory shows that the success of its constructions will require considerable cooperation from the world's causal structure. For example, consider the definition of simultaneity as absence of causal interaction. The definition alone cannot guarantee that the resulting slices are acausal and mutually exclusive; yet if two slices were to share an event, the account of temporal order fails to be unique. Consider the unfortunate situation depicted in Figure 2.0. The instants $I(e_1)$ and $I(e_3)$ intersect, yet we may suppose that, nevertheless, they are

both acausal world slices. But now consider the instant $I(e_2)$, not depicted above. Clearly both e_1 and e_3 are in $I(e_2)$, yet e_1 and e_3 are causally connectible. Thus $I(e_2)$ is not an acausal slice, contrary to the intention of the causal theory.

Leibniz was aware of this problem and explicitly set down a postulate which handled the difficulty. However, this and similar questions are better dealt with after the theory has been formulated more precisely.

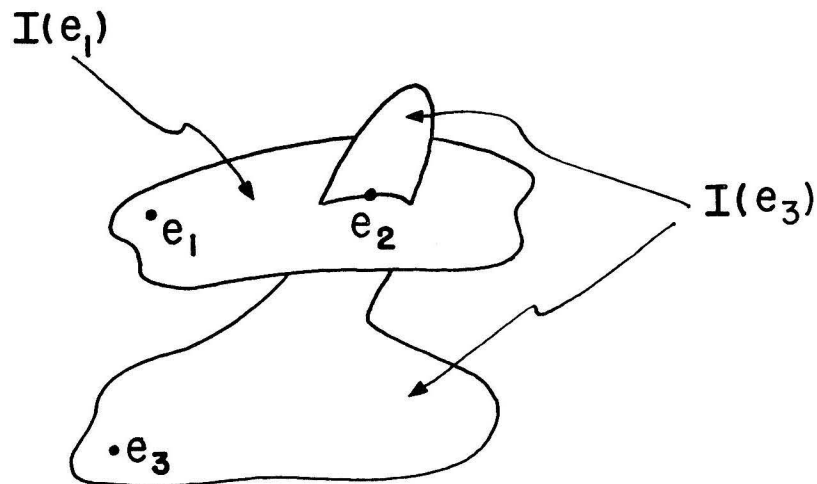


Figure 2.0

For working purposes let us use the relation of causal precedence, and read " $e_1 CP e_2$ " as "event e_1 causally precedes event e_2 ." Since there is no intention here to build into the relation *CP* a *direction* of causal order, "*CP*" could just as well be read "causally succeeds" throughout. This awkwardness might be avoided in a "purer" approach by beginning with a three-place relation of 'causal betweenness,' and then choosing an arbitrary pair of causally connected events to orient the event-manifold, and define *CP* accordingly.¹

To begin with, let us assume that the relation of causal precedence is irreflexive, asymmetrical, and transitive, i.e.,

Postulate 1. For all events e_1, e_2, e_3 :

- (1) not $e_1 CP e_1$,
- (2) if $e_1 CP e_2$, then not $e_2 CP e_1$, and
- (3) if $e_1 CP e_2$, and $e_2 CP e_3$, then $e_1 CP e_3$.

We may now define the relation of causal interaction or *causal connectibility*² ($e_1 \gamma e_2$) in the natural way:

Definition 2.0. $e_1 \gamma e_2$ iff $e_1 CP e_2$ or $e_2 CP e_1$.

The following proposition now follows easily:

Proposition 2.1. For all events e_1 , e_2 , and e_3 :

- (1) not ($e_1 \gamma e_1$)
- (2) if $e_1 \gamma e_2$, then $e_2 \gamma e_1$.

The Leibnizian definition of simultaneity is now given by:

Definition 2.2. e_1 Simul e_2 iff not ($e_1 \gamma e_2$).

Instants (of Time) are now event slices all of whose members are simultaneous with some event or other.

Definition 2.3. Let U be the set of all events. Then A is an instant iff $A \subseteq U$, and for some event e_0 in U , $A = \{e \mid e \text{ Simul } e_0\}$.

Assuming the success of these definitions, the set of all instants may be ordered as follows:

Definition 2.4. Let I_1 and I_2 be instants. Then I_1 temporally precedes I_2 ($I_1 < T I_2$) iff there are events $e_0 \in I_1$ and $e'_0 \in I_2$, such that $e_0 CP e'_0$.³

For this account to succeed, simultaneity must be an equivalence relation on the class (U) of all events. Proposition 2.1 and definition 2.2 yield reflexivity and symmetry for simultaneity at once; i.e., we have

Proposition 2.5. For all events e_1 and e_2 ,

- (1) e_1 Simul e_1 ,
- (2) if e_1 Simul e_2 , then e_2 Simul e_1 .

However, the transitivity of simultaneity does not yet follow, for we need assurance that if e_1 and e_2 are simultaneous, there is no other event e_3 , causally connected to e_1 but *not* causally connected to e_2 . As Figure 2.0 has shown, if this were to occur, the instant containing e_2 would also contain causally connected events (e_1 and e_3). Instants would not then be acausal sets, as intended. Thus we need to postulate that such a situation never arises:

Postulate II. (Leibniz postulate). If e_1 and e_2 are simultaneous and e_1 is causally connected with e_3 , then e_2 is also causally connected with e_3 .⁴

It now follows easily that Simul is an equivalence relation on U , i.e., we have:

Proposition 2.6. If e_1 Simul e_2 and e_2 Simul e_3 , then e_1 Simul e_3 . Simul is an equivalence relation on U .

Proof: Suppose e_1 Simul e_2 , e_2 Simul e_3 , yet not e_1 Simul e_3 . Then $e_1 \gamma e_3$, and by postulate II, $e_2 \gamma e_3$, thus contradicting e_2 Simul e_3 . Hence we must have that e_1 Simul e_3 . That Simul is an equivalence relation now follows from proposition 2.5. Done.

As a consequence, we also have justification for the above definition (2.4) of temporal precedence.

Proposition 2.7. Let I_1 and I_2 be nonempty instants. Then $I_1 < T I_2$ iff for all $e_1 \in I_1$, $e_2 \in I_2$, $e_1 CP e_2$.

Hence instants are acausal sets (i.e., no two events in an instant are causally connectible), and the order of instants has been defined consistently.

Clearly the success of the theory presupposes the correctness of the Leibniz postulate; here some further clarification is needed. To begin with, the postulate seems to be too strong. Suppose that events e_1 and e_2 are distant yet simultaneous. Will there not be, in some cases at least, an event e_3 causally connected with e_2 , say $e_2 CP e_3$, yet completely uninfluenced by the remote happening e_1 ? Not that some causal chain from e_1 *could not* terminate at e_3 ; but *do* such chains always so terminate, as a matter of actual fact?

Such a commitment may be avoided by construing our primitive causal relations modally, i.e., 'causally connected' becomes 'causally connectible,' etc.; we might even interpret the relation of 'causal connectibility' as a relation between events that is physically on a par with 'causally connected,' only more general.⁵ In any case, any such weaker interpretation should be construed so as to have causal connectedness entail causal connectibility, but not conversely. Postulates I and II, under such an interpretation, would, a priori, stand a better chance of being true.⁶

Of course, the most serious objection to the *truth* of the Leibniz postulate is that (even taken modally) it fails in the space-times of special and general relativity. In special relativity we have instead:

*Postulate II** (Einstein postulate). For any distinct events e_1 and e_2 such that e_1 Simul e_2 , there exists an event e_3 such that e_3 is causally connectible with e_2 , yet e_1 Simul e_3 .

At first sight, it seems that if this postulate were true, the result would be

the immediate demise of the causal theory. For while the Leibniz postulate (together with postulate I) guarantees that simultaneity will be an equivalence relation, the Einstein postulate virtually guarantees that it will not; since, according to this postulate, the only acausal instants are unit sets. Hence if we go on to postulate (as we must in any case) the existence of at least two simultaneous events, the Einstein postulate guarantees that an instant exists which is not acausal. However, it will soon be clear that the overthrow of the Leibniz postulate, rather than jeopardizing the causal theory, is a *necessary* condition for its *complete* success. In order to see this clearly, let us take up the account of the classical causal theory from where we left off.

So far, instants and a temporal order have been defined. However, as it stands, the theory of temporal order is incomplete. Are the causal ordering and its derivative temporal ordering discrete, dense, continuous, or of some other order type? We have not yet laid down any postulates to settle this question. However, their formulation in causal terms alone presents no difficulty. Thus we need add only the following postulate:

Postulate III. (Causal continuity).

- (1) For any event e there are events e_1 and e_2 such that $e_1 CP e CP e_2$.
(No first or last event.)
- (2) For any events e_1, e_2 such that $e_1 CP e_2$, there is an event e such that $e_1 CP e CP e_2$. (Density.)
- (3) For any nonempty sets of events A and B which are such that for any $e_1 \in A, e_2 \in B$, we have $e_1 CP e_2$, there exists an event e_0 such that if $e_1 \in A, e_2 \in B, e_1 \neq e_0 \neq e_2$, then $e_1 CP e_0 CP e_2$. (Continuity; cut Axiom, cf. Tarski, 1961.) Intuitively, this axiom states that if A and B are any sets of events such that the elements in A causally precede every element in B , then there is some event (e_0 , say) which separates A and B . The axiom fails for the rational numbers, as $A = \{x \mid x^2 < 2\}$ and $B = \{x \mid x^2 > 2\}$ show, since the required separator e_0 would be $\sqrt{2}$ —not a rational.
- (4) There is a denumerable set $K, K \subseteq U$, such that for any two events e_1, e_2 , where $e_1 CP e_2$, there is an event e in K such that $e_1 CP e CP e_2$. (Existence of a denumerable dense subset.)

It is now easy to show that the temporal order relation $<T$ on the set (Time) of all instants yields a structure ((Time), $<T$) which is isomorphic to the real numbers in their standard ordering.⁷ In other words, instants,

ordered by temporal precedence, have the same type of order as the reals in their standard ordering.

Still, the construction of a causal theory of space and time has barely begun. We must now go on to define: (1) space; (2) spatial topology and order; (3) the geometry or metric of space; and (4) the metric of time. First, the problem of space.

It was mentioned earlier that Leibniz spoke of space as the order of coexisting, i.e., simultaneous, events. It would be more correct to say that Leibniz is here characterizing spatial *order*, rather than space itself. An ordered *instant* is not (a) space; a space is a set of points that persist *through* the set of instants, and so is best construed as just the set of such persisting "points." Once space has been constructed, then we may turn to the question of the order of these points, and indeed might well attempt to base this order upon some relation between events within instants. The problem then comes down to defining points, and this just amounts to specifying for any given event e_0 , exactly what other event in any chosen instant occupies the *same* spatial point or place. Thus in Figure 2.1, let $I(e_0)$ be the instant that contains e_0 . Let I' be some other instant. What event e_0' in I' is at the same place as e_0 ? Once we know this, we may define a (spatial) point as the largest class of events all of which are at the same place. What we need then is a function that yields a different "fibre" (see Figure 2.1) through each event in a given acausal slice, with no two fibres intersecting, and each fibre meeting each instant in exactly one event. Each fibre is a spatial point; a space is the set of all fibres. Thus the problem of space, within the causal theory, is that of defining either a unique fibration of instants, or at least a distinguished class of such fibrations.

The Leibniz postulate precludes a nonarbitrary solution to this problem. Consider again the slices $I(e_0)$ and I' . Suppose that we assign e_0' and e_0 to the same spatial point or fibre as in the figure, rather than assigning, say, event e_1 in I' to the same spatial location as e_0 . But I' is a causally definable equivalence class (by the Leibniz postulate), so e_1 must sustain the *same* causal relations to e_0 as does e_0' . Thus a nonarbitrary association of e_0' and e_0 cannot be made on causal grounds alone; a fortiori, the same is true for any fibration of U . Hence the space problem cannot be solved utilizing only the resources of classical causality.

Typically, extra-causal aid is sought from point-particles and their dynamics. If it were physically possible to distinguish a privileged set of

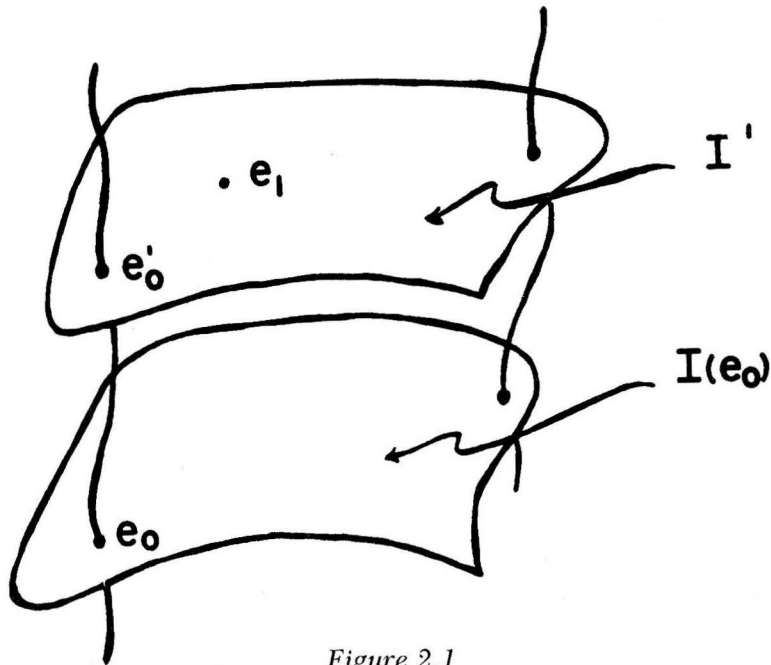


Figure 2.1

noncolliding particles, their world-lines might then be taken as our fibres or points, thereby defining a unique (absolute) space. A weaker approach is to distinguish a *family* of such sets of particles dynamically (inertial systems), with the result that we have a number of spaces, one for each system. Some of the awkwardness that results from having so many spaces on hand would then be reduced should it turn out that these spaces share the same (say Euclidean) geometry. But, once again, how are we to construct the geometry of space out of causality alone? If the metric is of the right sort (namely, Riemannian; cf. Hicks, 1971, pp. 70–71), we may use it to define the spatial topology; so let us consider only the spatial metric.

For simplicity's sake, consider the case in which we have constructed a unique space using a single dynamically distinguished set of point-particles (particles "at rest in absolute space"). The causal situation is completely characterized by the simultaneity slices, their temporal order, and the fact that any two events within a single particle (or fibre) are causally connectible. Once again, in order to make distance determinations possible, additional ontic resources are necessary. Typically, ex-

tended physical objects are now introduced, with a distinguished subset of these deemed "rigid." It is commonly believed, correctly, I think, that within the context of a classical causal theory and classical dynamics, there is an ineradicable element of physical convention which is involved in distinguishing rigid from nonrigid extended bodies; but this question is not at issue here.⁸

Similar considerations apply to the temporal metric. The continuity of the temporal order, as Riemann (1973) pointed out, precludes basing the temporal metric on the cardinality of the intervals between instants, for the simple reason that all intervals have the same cardinality. Once again, causality requires external aid, and typically, this is provided by *periodic* material processes (clocks). Just as in the case of rigid bodies, the conventions involved in distinguishing isochronous from irregular periodic process naturally and rightly emerge as crucial within such a construction.

The above arguments may be summarized as follows. The unaided classical causal theory provides no more than a theory of temporal order. To go beyond this, extra-causal considerations are necessary. Typically, dynamically free point-particles, rigid bodies, or periodic material processes are now introduced to construct space, its geometry, and the temporal metric. This weakness of classical causal theory derives in large part from the Leibniz postulate with its unique acausal slices of simultaneity. As a result, *any map of the universe U onto itself which preserves the temporal order relation will also be a mapping which preserves the total causal structure of U , i.e., such a map will be a causal automorphism of U .* This means, roughly, that we may stretch or shrink the classical time axis and "distort" the acausal slices at will, while preserving all of the causal relations of U . It thus follows (see theorem 2.10 below) that space, spatial geometry, and temporal metric cannot be constructed from classical causality alone. While the preceding arguments have appealed to this feature of classical causal structures in a more or less intuitive way, it is now time to proceed to demonstrate this rigorously.

Let us call any model of postulates I and II (the Leibniz postulate) a *Leibniz space-time*. In other words:

Definition 2.8. (U, CP) is a *Leibniz space-time* iff U is a nonempty set, CP

is a two-place relation on U , and for all events $e_1, e_2, e_3 \in U$,

- (1) $\langle e_1, e_1 \rangle \notin CP$ (irreflexivity),
- (2) If $\langle e_1, e_2 \rangle \in CP$, then $\langle e_2, e_1 \rangle \notin CP$ (asymmetry),
- (3) If $\langle e_1, e_2 \rangle \in CP$ and $\langle e_2, e_3 \rangle \in CP$, then $\langle e_1, e_3 \rangle \in CP$ (transitivity),

John A. Winnie

- (4) If $\langle e_1, e_2 \rangle, \langle e_2, e_1 \rangle \notin CP$ and $\langle e_2, e_3 \rangle \in CP$ or $\langle e_3, e_2 \rangle \in CP$, then either $\langle e_1, e_3 \rangle \in CP$ or $\langle e_3, e_1 \rangle \in CP$ (Leibniz postulate).

A causal automorphism of a Leibniz space-time is just a one-one map of its universe U onto itself which preserves the causality relation CP ; that is:

Definition 2.9. Let (U, CP) be a Leibniz space-time. Then h is a causal automorphism of (U, CP) iff h is a bijection: $U \rightarrow U$ such that for all $e_1, e_2 \in U, \langle e_1, e_2 \rangle \in CP$ iff $\langle h(e_1), h(e_2) \rangle \in CP$.

The causal automorphisms are the symmetries of a causal space; i.e., they are the ways of transforming it that do not alter its causality relation.⁹ What we want to show now is that any mapping that merely preserves the order of instants in a Leibniz space-time is a causal automorphism (the converse is trivial). This then is the root of the classical causal theory's inability to go beyond an account of temporal order *no matter how that theory is extended*. Here is the theorem and its simple proof.

Theorem 2.10. Let (U, CP) be a Leibniz space-time and $\langle T$ the temporal precedence relation defined as in definition 2.4 above. Then h is a causal automorphism of (U, CP) iff h is a bijection of U onto itself that preserves $\langle T$, i.e., for any instants I_1, I_2 (as defined in definition 2.3 above), $I_1 \langle T I_2$ iff $h[I_1] \langle T h[I_2]$.

Proof: The "only if" part is trivial. So suppose $I_1 \langle T I_2$ iff $h[I_1] \langle T h[I_2]$, where $h: U \rightarrow U$ is a bijection.

(1) Suppose $e_1 CP e_2$. Let $I(e_1)$ and $I(e_2)$ be instants containing e_1 and e_2 , respectively. By definition 2.4, $I(e_1) \langle T I(e_2)$, so $h[I(e_1)] \langle T h[I(e_2)]$. But $h(e_1) \in h[I(e_1)]$ and $h(e_2) \in h[I(e_2)]$. Hence by proposition 2.7, $h(e_1) CP h(e_2)$.

(2) Suppose $h(e_1) CP h(e_2)$. Let $I(e_1)$ and $I(e_2)$ be as in (1). Now $h(e_1) \in h[I(e_1)]$ and $h(e_2) \in h[I(e_2)]$. Since (by proposition 2.6) instants are equivalence classes, we must have $h[I(e_1)] = I[h(e_1)]$ and $h[I(e_2)] = I[h(e_2)]$. By definition 2.4, $I(h(e_1)) \langle T I(h(e_2))$, so we must have $h[I(e_1)] \langle T h[I(e_2)]$. But then $I(e_1) \langle T I(e_2)$, so by proposition 2.7, $e_1 CP e_2$. Thus $e_1 CP e_2$ iff $h(e_1) CP h(e_2)$, so h is a causal automorphism of $\langle U, CP \rangle$. Done.

One example of an application of this result will serve for our purposes. Let R^n be an n -place relation holding only between events within an instant and defined solely in terms of the causal relation CP . (For the sake of fixing ideas, consider R^n to be a four-place candidate for spatial congruence, the equidistance of e_1, e_2 and e_3, e_4 .) Since R^n is assumed definable

in terms of CP , any mapping that preserves CP must also preserve R^n . Let h_0 be any causal automorphism which preserves instants. Consider h_0 restricted to the instant $I_j(h_0/I_j)$. Tamper with h_0/I_j in any way you like as long as you leave the result $h_0^*/I_j: I_j \rightarrow I_j$. Clearly h_0^* preserves temporal precedence ($\langle T$) if h_0 does. Hence *any* h_0^* preserves R^n as well. So if $R^n(a_1, \dots, a_n)$ holds in I_j for distinct a_i , $R^n(b_1, \dots, b_n)$ must hold for *all* distinct b_i in I_j . Defining a spatial congruence relation is clearly impossible: a Leibniz space-time has far too many automorphisms.

The theorem upon which these results are based does not involve the assumption of temporal continuity (postulate III above) in its proof; in other words, the definition of a Leibniz space-time does not demand the temporal continuity of such spaces. Were we to go on now to postulate a *discrete* temporal order, a temporal metric would be obtainable by counting instants. Obviously, the (time) order-preserving maps would preserve this counting metric as well. (Clearly, a causally defined spatial metric is still impossible.)

However, the addition of the temporal continuity postulate leads immediately to the causal undefinability of a temporal metric. For the betweenness-preserving maps of the real line are just its topological homeomorphisms; so too then for instants and their order. Thus in Leibniz space-times with temporal continuity, causal structure is preserved by *any (topological) homeomorphism* of the set of instants. As in the case of space, no nontrivial temporal metric exists which is invariant under such a large set of transformations.

These results show that, for *Leibniz* space-times, temporal continuity precludes a causally definable temporal metric. But it should not be thought that continuity and homogeneity *alone* are the source of this metrical impotence. Rather, it is the continuity of time *taken together with the Leibniz postulate* that permits the proof of these results. As will become clear presently, continuity and homogeneity are properties of Minkowski causal spaces as well, but by virtue of the Leibniz postulate failing in such spaces and the Einstein postulate (postulate II* above) holding in its stead, the above result (theorem 2.10) fails to obtain. Indeed, according to a theorem of Zeeman (1964), any causal automorphism of Minkowski space-time maps a pair of congruent space-time intervals onto another pair of congruent intervals, although all space-time intervals, considered as sets of point-events, are continuous and have the same cardinality.¹⁰

III. Foundations of the Causal Structure of Minkowski Space-time¹¹

The construction of Leibniz space-time began with the two-place relation of causal precedence. In this section we shall see that the causal structure of Minkowski space-time may be derived from the *symmetrical* two-placed relation of causal connectibility.¹² Causal betweenness may then be defined in terms of causal connectibility, the space-time may be oriented, and the standard asymmetrical relations of causal and chronological precedence may be constructed easily. Along the way, it needs to be shown that the notions so defined do coincide with their intended counterparts within a standard interval formulation of the geometry of Minkowski space-time. Hence we shall need some such formulation to use as a reference throughout. The following definition will suffice for the present.

Let M be a nonempty set, and I^2 be a function from $M \times M$ into the real numbers. The pair (M, I^2) is said to be a *Minkowski space-time* if and only if there is a one-one function ϕ from M onto R^4 in which the interval I^2 is given by:

$$I^2(e_1, e_2) = -(\Delta x^0(e_1, e_2))^2 + \sum_k (\Delta x^k(e_1, e_2))^2, k = 1, 2, 3$$

where $\Delta x^i(e_1, e_2)$ is $(x^i(e_1) - x^i(e_2))$, and the functions x^i are the coordinate functions of ϕ (i.e., $x^i \equiv u^i \circ \phi$, where $u^i(a_1, \dots, a_n) \equiv a_i$). The coordinate functions $\{x^i\}$ derived from ϕ will be called a Lorentz coordinate system; x^0 is its time function. Frequently the function x^0 will be denoted by " t ," while " x ," " y ," and " z " will be used for x^1 , x^2 , and x^3 , respectively. (A signature of $(-, +, +, +)$ is used throughout.) The topology of M is that of R^4 and is induced by the function ϕ ; in other words, if $A \subseteq M$, then A is open in Minkowski space-time just in case its image, $\phi[A]$, is open in R^4 with its standard topology.

As our single causal relation we now take the relation (γ) of causal connectibility. Intuitively, event e_1 is causally connectible to event e_2 just in case a signal (massive or massless) can be sent from e_1 to e_2 , or conversely. In terms of (M, I^2) , $e_1 \gamma e_2$ just in case $I^2(e_1, e_2) \leq 0$. Note that, under this interpretation, the relation γ is both symmetrical and reflexive. However, γ is *not* transitive, since for any two events e_1 and e_3 which are simultaneous (not $e_1 \gamma e_3$), there will always be an event e_2 in their common causal futures.

It is important to realize that the above characterization of causal connectibility is not intended to be a *definition* of this relation. The present version of the causal theory takes γ as its only descriptive primitive relation. An axiomatic account, such as Robb's (1914), would now proceed to lay down a number of postulates for γ , each of which would be provably true under the above interpretation. However, an axiomatization of γ is not needed to show the adequacy of the causal theory of Minkowski space-time. For this purpose, we need only show that all geometrical notions in Minkowski space-time may be built up by explicit definitions from the causal connectibility relation. In order to do this we proceed on parallel "tracks." Having paired γ with its intended Minkowski interpretation, we may now consider another standard relation R on Minkowski space-time and propose $(--\gamma--)$ as a construction of R . The proof of this construction's adequacy consists in showing that if γ has its intended interpretation, then this is true of $(--\gamma--)$ as well. We continue in this way until we have provided constructions for a set of standard geometrical relations on Minkowski space-time which we know determine its geometry.

The following construction provides an example of this procedure. The intention is to construct the 'causal betweenness' relation from causal connectibility. The intended interpretation of " $CB e_1 e_2 e_3$ " is that the relation holds just in case there is a nonspacelike parameterized curve in M through e_1 , e_2 , and e_3 such that the parameter λ_2 of e_2 (i.e., $f(\lambda_2) = e_2$) is between that of e_1 and e_3 (i.e., $\lambda_1 \leq \lambda_2 \leq \lambda_3$ or $\lambda_3 \leq \lambda_2 \leq \lambda_1$). An equivalent condition (in Minkowski space-time) is that e_1 and e_3 are causally connectible and e_2 is in the closure of the Alexandroff interval¹³ of e_1 and e_3 . Obviously, other equivalent characterizations of causal betweenness are possible, and we may use our general knowledge of Minkowski space-time to establish their equivalence to either of the above conditions. From the standpoint of the causal theory, the interesting fact is that causal betweenness may be characterized solely in terms of causal connectibility, as the following definitions and the justifying theorem show.

First of all, consider the set of all events $J(e)$ causally connectible with a given event e . This is just the set of all events e_1 such that $I^2(e_1, e) \leq 0$ and so is definable as

Definition 3.0. $J(e) = \text{df. } \{e_1 \mid e_1 \gamma e\}$.

Clearly this definition is accurate, so a justifying theorem will not be

given. However, a proof of the adequacy of the following definition of *causal betweenness* is provided below.

Definition 3.1. $CB e_1 e_2 e_3 = \text{df. } e_1 \gamma e_3 \text{ and } J(e_2) \subseteq J(e_1) \cup J(e_3).$

Proposition 3.2. $CB e_1 e_2 e_3$ if and only if e_2 is causally between e_1 and e_3 .

Proof: (A). Suppose e_2 is causally between e_1 and e_3 . Then e_1 is causally connectible to e_3 , and if $e_1 = e_3$, $e_2 = e_1 = e_3$. Hence, trivially, $J(e_2) \subseteq J(e_1) \cup J(e_3)$. Suppose $e_1 \neq e_3$, with e_2 causally between e_1 and e_3 . Let e^* be causally connectible with e_2 (i.e., $e^* \in J(e_2)$). Then there is a causal curve from e_2 to e^* , and a causal curve from either e_1 or e_3 to e_2 .¹⁴ Hence there is a causal curve from either e_1 or e_3 to e , so $e \in J(e_1) \cup J(e_2)$.

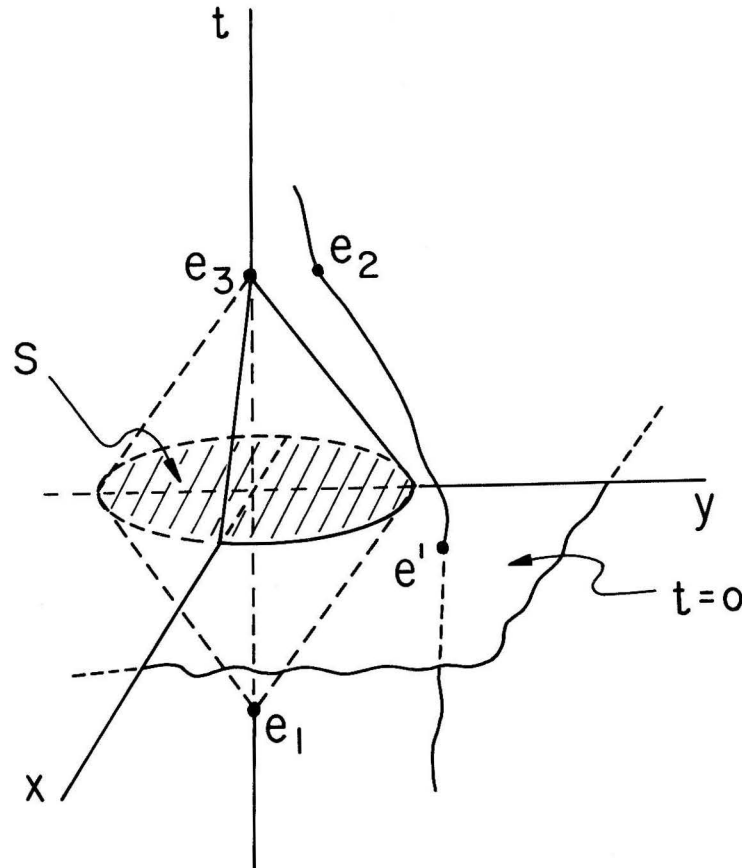


Figure 3.0

(B). Suppose e_2 is not causally between e_1 and e_3 , with $I^2(e_1, e_3) < 0$. Let $\{x^i\}$ be a Lorentz coordinate system in which $t(e_1) = -1$, $t(e_3) = +1$ and $x^k(e_1) = x^k(e_3) = 0$ ($k = 1, 2, 3$). Then the intersection of the $t = 0$ hypersurface with the set of all events causally between e_1 and e_2 is the set of all events S with coordinates $t = 0$ and $\sum (x^k)^2 \leq 1$ (see Figure 3.0). Those events causally between e_1 and e_3 are just the domain of dependence of S , i.e., for all and only such events e^* will it be the case that every endless timelike curve through e^* meets S (Geroch, 1970). Hence if e_2 is not causally between e_1 and e_3 , there is an endless timelike curve through e_2 which does not meet S , and thus does not meet hypersurface $t = 0$ outside S at some event e' simultaneous with e_1 and e_3 (i.e., not $e' \gamma e_1$ and not $e' \gamma e_3$). Hence $e' \in J(e_2)$ but $e' \notin J(e_1) \cup J(e_3)$.

When $I^2(e_1, e_3) = 0$, the set S' of all events causally between e_1 and e_3 is just the corresponding segment of the null line determined by e_1 and e_3 . Thus the domain of dependence of S' is just S' itself (Geroch, 1970), and the preceding argument may be applied once more. Done.

Causal connectibility and betweenness also allow us to distinguish the boundary of the causal space $J(e)$ from its interior. We begin by defining the appropriate connectibility relations. The corresponding betweenness relations then follow at once.

Events e_1 and e_2 are said to be *light-connectible* just in case $I^2(e_1, e_2) = 0$. The light-connectibility of two events may be defined in terms of causal connectibility as follows:

Definition 3.3. $e_1 \lambda e_2 = \text{df. } e_1 = e_2, \text{ or } e_1 \neq e_2, e_1 \gamma e_2, \text{ and } LB = \{e \mid CBe_1 e e_2\}$ is causally connectible, i.e., for any two events, e, e' , in LB , $e \gamma e'$.

Proposition 3.4. $e_1 \lambda e_2$ iff $I^2(e_1, e_2) = 0$.

Proof: The case of $e_1 = e_2$ is immediate. So suppose $e_1 \neq e_2$. In Minkowski space-time, the null line containing two light-connectible events e_1 and e_2 is a causally connectible set. Hence the same is true of the set of all events causally between e_1 and e_2 , since they all lie on this line. However, if e_1 and e_2 are causally connectible but have timelike separation, the intersection of the causal future of one with the causal past of the other will contain at least two events which are not causally connectible. Done.

Events e_1 and e_2 are chronologically connectible just in case they have

timelike separation, i.e., $I^2(e_1, e_2) < 0$. Clearly the following definition succeeds:

Definition 3.5. $e_1 \tau e_2 = \text{df. } e_1 \gamma e_2 \text{ and not } e_1 \lambda e_2$.

Notice that since $e \lambda e$, chronological connectibility is defined so as to be irreflexive (not $e \tau e$), while causal and light connectibility are both reflexive.

With timelike and lightlike connectibility on hand, the causal betweenness relation may be refined further to yield chronological betweenness and null betweenness, the latter being the order of events along null lines. The intent and adequacy of the definitions below are both obvious, so justifying theorems are left as exercises.

Definition 3.6. (a) $TB e_1 e_2 e_3 = \text{df. } CB e_1 e_2 e_3 \text{ and } e_1 \tau e_2, e_3$.
 (b) $LB e_1 e_2 e_3 = \text{df. } CB e_1 e_2 e_3 \text{ and } e_1 \lambda e_3$.

The above constructions show that, although we start with causal connectibility alone, we may nevertheless distinguish events that are "inside" the causal space of e ($\{e_1 \mid e_1 \tau e\}$) from those that are on its "surface" ($\{e_1 \mid e_1 \lambda e\}$). It should be noted, however, that these constructions rely upon global features of Minkowski space-time and will fail to be adequate in Lorentz space-times in general. For example, in the two-dimensional Minkowski space-time depicted in Figure 3.1, if segment K is removed, definition 3.1 of causal betweenness fails, since the causal space of e_2 is a subset of the causal space of e_1 , yet e_2 is not causally between e_1 and e_3 .

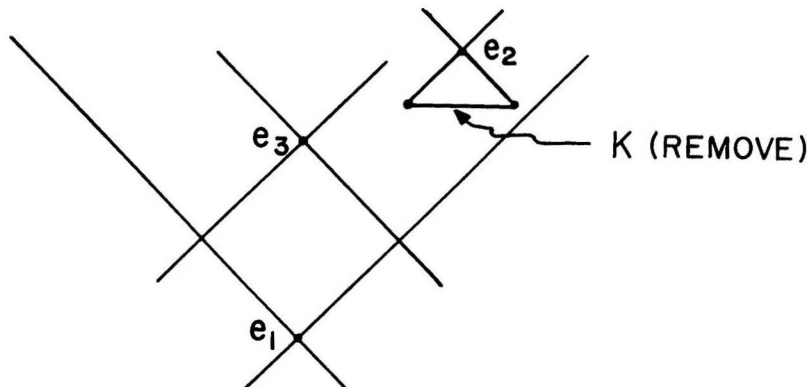


Figure 3.1

At this point, it is possible to proceed at once to the development of synthetic causal geometry. However, it is technically convenient to work with asymmetrical causal relations such as causal "precedence," and much contemporary causal research—also Robb—employs such a relation from the outset. While the use of such relations does not presuppose a direction (a privileged future) to causality, it does assume that space-time is *causally orientable*, that is, an arbitrary but consistently assigned causal direction may be selected. Hence, since Minkowski space-time is causally orientable, the question might just as well be faced now so that we may freely use the result in later developments.

The causal space of an event e breaks up into two components: its causal "future" J^+ and its causal "past" J^- . The event e is here considered to be a member of both J^+ and J^- . Now one cannot speak of a "future" or "past" without having first oriented the space-time. Nevertheless, without orientation, one can say that one or the other component of the causal space J of e is the component in which a given event e_1 resides. This may be accomplished using the causal betweenness relation as follows:

Definition 3.7. Where $e_2 \gamma e_1$, and $e_2 \neq e_1$, the e_2 -component of e_1 is just the set of all events e in the causal space of e_1 such that e_1 is not causally between e and e_2 , together with e_1 , i.e.,

$$\text{Comp } e_2(e_1) = \text{df. } \{e \mid e \in J(e_1) \ \& \ \text{not } CB ee_1e_2\} \cup \{e_1\}.$$

Proposition 3.8. Let $\{x^\alpha\}$ be a Lorentz coordinate system for a Minkowski space $\mathfrak{M} = (M, I^2)$. Let e_1 be an event in M , with e_2 in the causal future (past) of e_1 . (Here $\{x^\alpha\}$, as usual, is used to orient \mathfrak{M} .) Then for any event e , e is in the causal future (past) of e_1 if and only if $e \in \text{Comp } e_2(e_1)$.

Proof: If $e = e_1$, the result is immediate. So assume $e \neq e_1$. Suppose e is in the causal future of e_1 , yet $CB ee_1e_2$. Then there would be a causal curve from e_1 to e and from e to e_1 —which, in Minkowski space-time, is impossible. Next, suppose that $e \in \text{Comp } e_2(e_1)$, i.e., $e \gamma e_1$ and not $CB ee_1e_2$. Since $e \gamma e_1$, there is either (a) a causal curve from e to e_1 or (b) a causal curve from e_1 to e . If (a), since there is a causal curve from e_1 to e_2 , there is a causal curve from e to e_1 to e_2 , so we would have $CB ee_1e_2$, contrary to our assumption. Hence (b) must hold, so e is in the causal future of e_1 . Done.

We now *causally orient* Minkowski space-time by selecting an ordered pair of events $\mathfrak{C}_M \equiv \langle e_i, e_j \rangle$ which are causally connectible ($e_i \gamma e_j$), but not identical. Such an event-pair will be called an *orientation* of \mathfrak{M} , and

$\mathfrak{M}' = (M, I^2, \mathcal{O}_{ij})$ is called an *oriented Minkowski space-time*. Let F_{ij} be the e_j component of e_i , with P_{ij} the other component of e_i (i.e., those events in $J(e_i)$, which fail to be in the e_j component of e_i , together with e_i itself). We now define the *causal precedence* relation as follows:

Definition 3.9. Let events e_1 and e_2 be causally connectible. Then e_1 *causally precedes* e_2 with respect to orientation $\mathcal{O}_{ij} = \langle e_i, e_j \rangle$, i.e., $e_1 < e_2$ (reference to \mathcal{O}_{ij} omitted) iff (see Figure 3.2):

- (1) $e_1 = e_i$ and $e_2 \in F_{ij}$, or $e_1 \neq e_i$ and either
- (2) not $e_1 \gamma e_i$ and $\text{Comp } e_2(e_1) \cap F_{ij} \neq \{ \}$, or
- (3) $e_1 \in F_{ij}$ and $\text{Comp } e_2(e_1) \subseteq F_{ij}$, or
- (4) $e_1 \in P_{ij}$ and $\text{Comp } e_2(e_1) \not\subseteq P_{ij}$.

We show the adequacy of this definition as follows. Let $\mathfrak{M}' = (M, I^2, \mathcal{O}_{ij})$ be an oriented Minkowski space-time with orientation $\mathcal{O}_{ij} = \langle e_i, e_j \rangle$. Now let $K \equiv \{x^\alpha\}$ be a Lorentz coordinate system for \mathfrak{M}' . If $\{x^\alpha\}$ is such that $t(e_i) < t(e_j)$, then $\{x^\alpha\}$ will be said to be *compatible* with \mathfrak{M}' ; otherwise K and \mathfrak{M}' will be called *incompatible*. Clearly, if \mathfrak{M}' and K are compatible, the function t makes F_{ij} the causal future of e_i , rather than its causal past. That this works out for all events in M is shown by the following proposition:

Proposition 3.10. Let $\mathfrak{M}' = (M, I^2, \mathcal{O}_{ij})$ be an oriented Minkowski

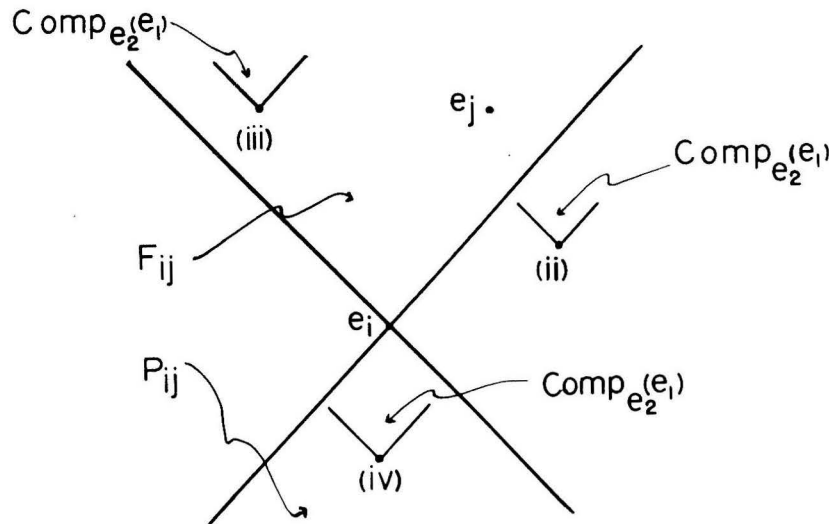


Figure 3.2

space-time, with $\{x^\alpha\}$ a Lorentz coordinate system compatible with \mathfrak{M}' . Then $e_1 < e_2$ iff e_1 is causally connectible with e_2 and $t(e_1) \leq t(e_2)$.

Proof: We consider the cases as in definition 3.9. Case (1): immediately follows from the adequacy of the definition of components. Case (2): the F_{ij} component of e_i must intersect some component of e_1 . If this were its past component, there would be a causal curve from e_i to e_1 , contradicting the assumption that not $e_i \gamma e_1$. Hence F_{ij} must intersect the future component of e_1 , which is the intended component. Case (3): in Minkowski space-time, when $e \neq e'$, e' is in the causal future of e iff the causal future of e' is a subset of the causal future of e . Case (4): similar to (3). Done.

Given an orientation \mathcal{O}_{ij} , we may now define the causal future of e ($J^+(e)$) as the set of all events which e causally precedes, and the causal past of e ($J^-(e)$) as the set of all events that causally precede e . That is,

- Definition 3.11. (a) $J^+(e) = \text{df. } \{e_1 \mid e < e_1\}$
- (b) $J^-(e) = \text{df. } \{e_1 \mid e_1 < e\}$

Chronological precedence (\ll) and precedence along a null line (\rightarrow)—sometimes called the *horismos* relation—are now definable using causal precedence and the appropriate connectibility relation in an obvious way.

- Definition 3.12. (a) $e_1 \ll e_2$ (e_1 chronologically precedes e_2) = df. $e_1 < e_2$ and $e_1 \tau e_2$
- (b) $e_1 \rightarrow e_2$ (horismos relation) = df. $e_1 < e_2$ and $e_1 \lambda e_2$.

These definitions provide the basic causal notions to be used throughout this account. Although these constructions are mere preliminaries to the construction of Minkowski geometry proper, it is easy to see that, even at this preliminary level, not all of the preceding constructions succeed in arbitrary Lorentz space-times. This tendency of Minkowski space-time to cooperate with causal constructions will become even more evident as we proceed to construct its linear and metrical structure causally. However, before examining the causal basis of Minkowski metrical geometry, there is an interesting preliminary matter to consider: the causal basis of the *topology* of Minkowski space-time.

IV. Causality and Space-time Topology

According to the causal theory, the entire geometry of Minkowski space-time may be derived from its causal structure. Not only space-time

metric or congruence, but the linear structure and topology of Minkowski space-time, must be definable in terms of causality. For some geometries (metric spaces), the problem reduces to defining the metric, since the linear structure and topology of such spaces is contained in their metric. However, this is not the case for the indefinite metrics of relativity theory. For these spaces, the connection between metric and topology is more indirect. Nevertheless, for many of these spaces, Minkowski space-time included, their causal structure permits a simple and direct definition of their topology. So while special relativity is the main concern here, the following causal construction of standard space-time topology succeeds in many of the space-times encountered in general relativity as well.

In the preceding section Minkowski space-time was defined as a pair (M, I^2) , where the semi-Riemannian metric I^2 is specified in terms of a bijective coordinate mapping ϕ from M to R^4 . However, the mapping ϕ was also used to define the standard R^4 topology on M by simply calling a subset N of M open just in case $\phi[N]$, its image, is open in the standard topology of R^4 . Thus the mapping ϕ serves two independent functions: it provides M with a "metrical" structure by specifying the interval I^2 , and, in addition, ϕ is used in a quite different way to induce the R^4 metric topology on M . Naturally, the question arises: what is the relation between these two structures on M ? In particular, can the topology of Minkowski space-time be derived from the space-time interval I^2 ?

The problem of the independence of the topological and metrical structures of space-time was clearly recognized by early writers on relativity such as Russell (1954) and, of course, Eddington:

. . . the statement that the world is four-dimensional contains an implicit reference to some ordering relation. This relation appears to be the *interval*, though I am not sure whether that alone suffices without some relation corresponding to *proximity*. It must be remembered that if the interval s between two events is small, the events are not necessarily near together in the ordinary sense. (1959, pp. 186–187.)

Eddington's *proximity* is just the local topological structure of space-time. Whereas Russell (1954, p. 56 ff.) held that space-time topology is interval-independent, Eddington did not rule out interval-dependence altogether, although he did recognize that the topology of space-time is not derivable from its interval in the standard way. As will become clear soon, Eddington's caution was justified. For the topology of many space-

times (surely all those Eddington and Russell were considering) is obtainable from their space-time metric tensor field, although it is the *causal* content of the metric tensor which is exploited.

To begin with, let us be clear about just what aspect of relativistic space-times causes the difficulty. Consider a set N and a function d from $N \times N$ into R_0^+ (the set of nonnegative real numbers). If d is such that:

- (1) $d(x, x) = 0$; (2) if $d(x, y) = 0$, then $x = y$;
 (3) $d(x, y) = d(y, x)$, and (4) $d(x, y) + d(y, z) \geq d(x, z)$,

then (N, d) is said to be a *metric space*. When (1), (3), and (4) hold, the result is called a *pseudo-metric space*. If we let N be the point-set of either a Euclidean space or any of the classical non-Euclidean spaces, and take d to be the corresponding distance function on these point-sets, then each of these geometries may be shown to be a metric space in the above sense.¹⁵ Furthermore, when a space (N, d) is a metric space, there is a natural way of defining a topology on that space: a subset A of N is open just in case, for every point x_0 in A , there is a *ball* $B_\epsilon(x_0) \equiv \{y \mid d(x_0, y) < \epsilon\}$, $\epsilon > 0$, containing x_0 , and at the same time wholly contained in A ($B_\epsilon(x_0) \subseteq A$). The fact that the open sets so defined indeed yield a topology on N is easy to prove, but the proof does require that the space (N, d) be a pseudo-metric space. For metric spaces, and thus for the classical geometries, their metric yields their standard topology when the above definition is used. More generally, it can be shown that for any Riemannian space with positive definite metric, the metric topology and manifold topology coincide (see Hicks, 1971, pp. 70–71).

However, Minkowski space-time is neither a metric nor a pseudo-metric space. The Minkowski interval yields distinct events with zero separation, and the triangle inequality fails. It is then easy to show that if "open" sets are defined in terms of balls as above, they do not determine *any* topology on M , much less the standard (R^4) topology. For the indefinite space-times of relativity theory, metrical nearness and topological nearness do not coincide. These facts appear to be the basis for Russell's conclusion that the topology of a space-time cannot be derived from its interval.

In retrospect, it is easy to see that this conclusion was too hasty. For although the *usual* definition of a topology in terms of balls given by the space-time interval will not do, it does not follow that some *other* definition in terms of the interval cannot succeed. Indeed, we now know that

for a wide class of relativistic space-times (Minkowski space-time included), another type of construction does yield their standard topology. Moreover, this definition utilizes only that content of the interval which relates to the causal structure of space-time. In brief, the standard topologies of Minkowski and other space-times may be derived solely from their causal structure. Such a derivation is the first step on the way to a full-fledged causal theory of Minkowski geometry.

Like any topology, a topology for Minkowski space-time may be defined by specifying just which sets are to constitute its basis. The causal theory demands that these basis sets be constructed out of causality relations on M . These chosen basis sets are usually called *Alexandroff intervals* and are causally defined as follows:

Definition 4.0. For any events e_1, e_2 in M , the *Alexandroff interval* of e_1 and e_2 , $Alex(e_1, e_2)$, is just $\{e \mid TB(e_1ee_2)\}$: the set of all events chronologically between e_1 and e_2 (see Figure 4.0).

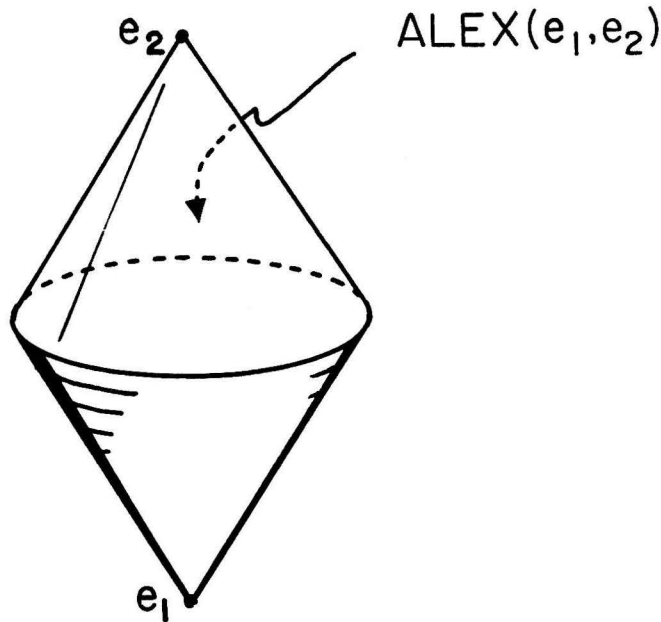


Figure 4.0

When e_1 and e_2 are chronologically connectible, then an Alexandroff interval consists of those events in the interior of the cone depicted in Figure 4.0; otherwise $Alex(e_1, e_2)$ is simply the empty set. It is now easy to show that the Alexandroff intervals do provide a basis for a topology on M , i.e., for any two such intervals A_1 and A_2 , there is another which is a subset of their intersection. A subset A of Minkowski space-time is now said to be open in the *Alexandroff topology* on M just in case every event e in A is an element of an Alexandroff interval which is itself a subset of A . Clearly in Minkowski space-time, any Euclidean 4-ball contains an Alexandroff interval, and conversely. Thus we have the result: *the Alexandroff topology and the standard E^4 -topology on M coincide* (Hawking and Ellis, 1953, p. 196 ff). Since an Alexandroff interval is defined causally, this straightforward result establishes the adequacy of a causal theory of the standard topology of Minkowski space-time.¹⁶

Thus Russell and Eddington's questions concerning the physical significance of the topology of space-time are answered by Alexandroff's simple construction, at least for the space-time of special relativity. Russell's conclusion that space-time topology is independent of its metric was not justified, although he saw quite clearly that the topology of space-time does not have the *customary* relation to the space-time metric, and also realized that the space-time interval has an intimate connection with causality (1954, chap. 35). However, Russell's work on the foundations of relativity theory was seriously flawed by his being misled by some of the accidental features of the interval, such as its assignment of zero distances to all events on the same light path.¹⁷ In general, Russell read the space-time interval too literally, and as a result never became clear about its relation to causal structure.

While the Alexandroff construction succeeds at once in Minkowski space-time, the situation in general relativity is not so straightforward. In some cases, the counterpart to the Alexandroff topology will be coarser than (have "fewer" open sets than) the manifold topology of general relativistic space-times. However, the two topologies do coincide in all Lorentz space-times that satisfy the condition of *strong causality*. This condition says, roughly, that every event e has arbitrarily small neighborhoods to which no causal curve returns once having left that neighborhood (Hawking and Ellis, 1973, p. 192ff). Since space-times that violate strong causality are in some sense "pathological," the physical significance of these exceptions to a causally definable topology is by no means clear at present.

V. Synthetic Causal Geometry

A. A. Robb appears to have been the first to suggest that the geometry of the space-time of special relativity can be developed on a purely causal basis. He was certainly the first to work out the details of this idea, and the result was a profound work of considerable magnitude and difficulty (1914).¹⁸ Given Robb's ambitious aims, and the conceptual economy of his foundations, the complexity of his constructions was inevitable. For Robb's account was axiomatic, and the causal theorist who wishes to axiomatize his theory cannot avoid this consequence of having but a single primitive relation at his disposal. Robb's choice was the relation $<$ of causal precedence introduced in the preceding section, and twenty-one postulates are laid down for this relation. Only after a few hundred theorems do coordinates appear, and with them the Minkowski metric in its standard form. On the other hand, the definition of Minkowski space-time in section III above used Lorentz coordinates to achieve its result with a single postulate. Nothing could be a better example of the power of coordinate methods to introduce in a single step a wealth of structure which we do not begin to understand!

Rather than attempt a detailed survey of Robb's constructions, this section will provide a more concise development of the affine geometry of Minkowski space-time. The transition from affine to metric geometry is developed in the following section, although a few of the most important of Robb's metrical constructions are sampled here. The philosophical interest of these constructions lies in the fact that they show how the causal theory solves the central problem of any relational theory: the construction of quantitative (or metrical) comparison out of a qualitative causal relation. As we have seen, most contemporary causal theorists (Robb excepted) have turned to extra-causal sources as the basis of congruence and metric, and furthermore, they have provided epistemological and ontological arguments that purport to demonstrate the necessity for introducing extra-causal considerations. The following developments show how the causality relation alone permits the construction of the affine (linear) structure and interval congruence in space-time. As a result, these same constructions show how Reichenbach's arguments for the necessity of coordinative definitions of congruence (1957, chap. 1) are circumvented by the causal theory. The same is true of arguments that appeal to the homogeneity and continuity of a manifold as demanding an

extrinsic basis for congruence in opposition to the intrinsic temporal order afforded by causal relatedness (as in Grünbaum (1969), sec. 3). Both types of argument fail as a result of having overlooked the possibility of a purely causal construction of spatial, temporal, and space-time congruence. The account of the preceding section shows that this underestimation of the resources of causality is understandable; nevertheless, the results of the succeeding sections show this to have been a serious philosophical error.

The constructions that follow outline three areas of space-time geometry: linearity, orthogonality, and congruence. The development of the linear (affine) structure of Minkowski space-time is a considerable simplification of Robb's approach; proofs of the adequacy of its definitions are reasonably direct, and are either cited or provided. The account of orthogonality and congruence follows Robb closely; however, it is incomplete, and the definitions are not accompanied by adequacy proofs. This is remedied in section VI. With few exceptions, the constructions are not *obviously* successful—in general, their adequacy depends upon the world's causal structure in a nontrivial way. The remarkable fact is that, in the case of special relativity, the right sort of structure is forthcoming.

We begin with the linear structure of Minkowski space-time, i.e., the construction of its straight (geodesic) lines, planes, parallelism, and hyperplanes.

Lightlike (null) lines are the easiest to construct. If events e_1 and e_2 are light-connectible, i.e., $e_1 \lambda e_2$, the *null line containing* e_1 and e_2 ($L(e_1, e_2)$) is just the set of all events e which are light-connectible to both e_1 and e_2 ; in general, a set is a null line just in case it is a set of events such that for some distinct events e_1, e_2 , this set is just $L(e_1, e_2)$.

Definition 5.0. When $e_1 \lambda e_2$, and $A \subseteq M$,

- (a) $L(e_1, e_2) = \text{df. } \{e \mid e \lambda e_1 \ \& \ e \lambda e_2\}$
- (b) A is a null line = df. for some distinct e_1, e_2 such that $e_1 \lambda e_2$, $A = L(e_1, e_2)$.

Although the adequacy of this definition is fairly obvious, a proof is provided for the sake of completeness.

Proposition 5.1. The null line L determined by two distinct events e_1 and e_2 with zero separation is just the set of events light-connectible to both e_1 and e_2 .

Proof: Choose a Lorentz frame in which the coordinates of e_1 and e_2 (0, 0,

0, 0) and (1, 1, 0, 0), respectively. Those events light-connectible with e_1 have coordinates satisfying:

$$(1) -(x^0)^2 + \sum (x^k)^2 = 0, \quad k = 1, 2, 3,$$

while those events light-connectible to e_2 satisfy

$$(2) -(x^0 - 1)^2 + (x^1 - 1)^2 + (x^2)^2 + (x^3)^2 = 0.$$

Subtracting (1) and (2) gives

$$(3) x^0 = x^1,$$

and substituting (3) in (1), we obtain

$$(4) x^2 = x^3 = 0.$$

Hence the points satisfying (1) and (2) are those of the form $(a, a, 0, 0)$, i.e., just those points on the null line L determined by e_1 and e_2 . Done.

The construction of spacelike lines is not so obvious. Robb's approach is to use null lines to construct planes of a special kind, which he then intersects to recover non-null lines. The following idea is due to Latzer (1972), and is much simpler.

Suppose that the separation of e_1 and e_2 is spacelike (i.e., not $e_1 \gamma e_2$). Then we say that e_3 is *S-collinear with e_1 and e_2* just in case e_3 has a spacelike separation to both e_1 and e_2 (not $e_3 \gamma e_1$ and not $e_3 \gamma e_2$), and the light cones of e_1, e_2 , and e_3 fail to have a common point of intersection (there is no event e such that $e \lambda e_1, e_2, e_3$). The line $S(e_1, e_2)$ determined by e_1 and e_2 is now just the set of all events S-collinear with e_1 and e_2 . That is,

Definition 5.2 (a) $Se_1e_2e_3 = \text{df.}$ (1) not $(e_1 \gamma e_2, e_3)$ & not $(e_2 \gamma e_3)$, and (2) there is no event e_4 such that $e_4 \gamma e_1, e_2, e_3$.

(b) A is a spacelike line = df. $A \subseteq M$ and for some events e_1 and e_2 , $A = \{e_3 \mid Se_1e_2e_3\}$.

The adequacy of this construction is shown by Latzer (1972, proposition 2.11).

Timelike lines may be constructed using the following approach. Let e_1 and e_2 be chronologically connectible ($e_1 \tau e_2$). Next, consider the intersection of their light cones; this is defined as just $LC(e_1, e_2) \equiv \{e \mid e \lambda e_1 \text{ and } e \lambda e_2\}$, and in a Lorentz coordinate system in which $x^k(e_1) = x^k(e_2) = 0$ ($k = 1, 2, 3$), $t(e_1) = -2$, $t(e_2) = 0$, this intersection will be a spacelike three-sphere defined by $(x^1)^2 + (x^2)^2 + (x^3)^2 = 1$ in the $t = -1$ hypersurface. (See Figure 5.0, one dimension suppressed.) Connect any two points in this sphere by a spacelike line. The set of all such lines yields a spacelike

hypersurface $H(e_1, e_2)$, which is just the $t = -1$ surface of Figure 5.0. This hypersurface now being fixed, any other event e_3 is said to be *temporally collinear with e_1 and e_2* ($T(e_1, e_2, e_3)$) just in case the hypersurface $H(e_3, e_2)$ (or $H(e_3, e_1)$) does not intersect the canonical hypersurface $H(e_1, e_2)$. In the figure, e_3 is thus collinear with e_1 and e_2 whereas e_4 is not. The following definition recapitulates this construction, and the theorem which follows establishes its accuracy.

Definition 5.3. Where $e_1 \tau e_2$:

(a) $LC(e_1, e_2) = \text{df.}$ $\{e \mid e \lambda e_1, e_2\}$

(b) $H(e_1, e_2) = \text{df.}$ $\{A \mid A \text{ is a spacelike line such that there are at least two events in } LC(e_1, e_2) \cap A\}$

(c) $T(e_1, e_2, e_3) = \text{df.}$ (1) $e_3 = e_1$ or $e_3 = e_2$ or (2) $e_3 \neq e_1, e_2$, and $e_2 \tau e_3$ & $H(e_1, e_2) \cap H(e_2, e_3) = \{ \}$.

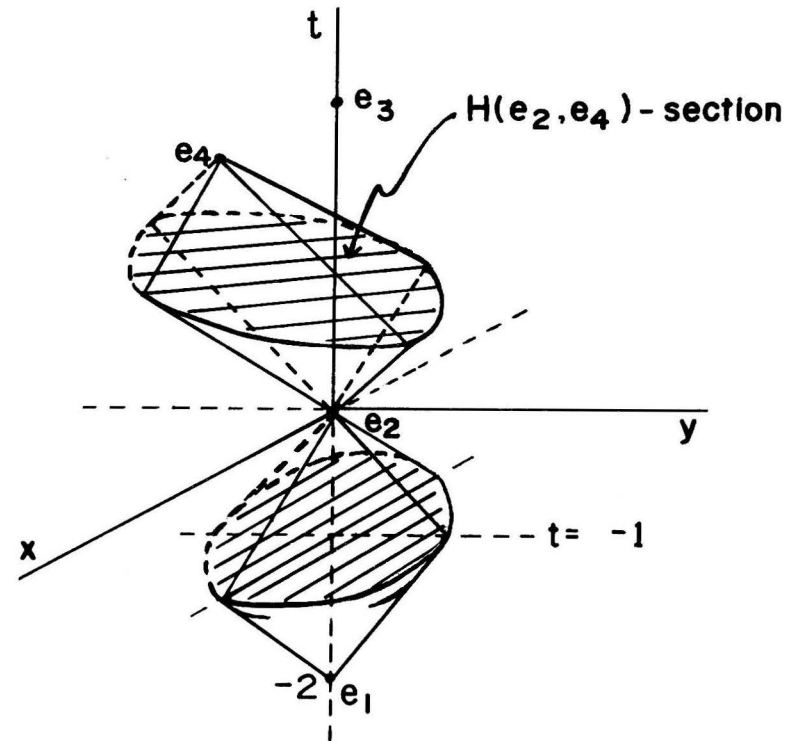


Figure 5.0

(d) A is a timelike line = df. $A \subseteq M$ and for some $e_1, e_2 \in M$, such that $e_1 \tau e_2$, $A = \{e_3 \mid T(e_1, e_2, e_3)\}$.

Proposition 5.4. If e_1 and e_2 have timelike separation, then e_3 is collinear with e_1 and e_2 iff $e_2 \tau e_3$ and $\text{Hyp}(e_1, e_2) \cap \text{Hyp}(e_2, e_3) = \{ \}$.

Proof: Let e_1 and e_2 have timelike separation, and e_3 be any event such that $e_2 \tau e_3$. Choose a Lorentz frame $K = \{x^\alpha\}$ in which: (1) $x^0(e_1) = -2$, $x^0(e_2) = 0$; and (2) e_3 is in (say) the first quadrant of the $x^0 - x^1$ plane, and thus has coordinates $(a_0, a_1, 0, 0)$ (see Figure 5.0). If e_3 is identical with either e_1 or e_2 , we are done. Otherwise, let K' be the Lorentz frame with relative speed $v = a_1/a_0$ to K in the positive direction along the x^1 -axis of K , K' being thus the frame in which e_3 lies along the $x^{0'}$ -axis. By the Lorentz transformations $x^{0'}(e_3) = ((a_0)^2 - (a_1)^2)^{1/2}$ and thus hypersurface $H(e_2, e_3)$ is given by

$$(1) \quad x^{0'} = \frac{x^{0'}(e_3)}{2} = \frac{((a_0)^2 - (a_1)^2)^{1/2}}{2},$$

in terms of the K' coordinates. Transforming back to K we obtain:

$$(2) \quad a_0 x^0 - a_1 x^1 = \frac{((a_0)^2 - (a_1)^2)}{2}$$

The $H(e_1, e_2)$ hypersurface is given by $x^0 = -1$, so the intersection of the two hypersurfaces is given by substituting in (2) to obtain

$$(3) \quad x^1 = \frac{((a_1)^2 - (a_0)^2 - 2a_0)}{2a_1}$$

which will have a solution just in case $a_1 \neq 0$, i.e., just in case e_3 fails to be on the x^0 -axis of K . Done.

Having constructed the straight lines of Minkowski geometry, we may now go on to define its other linear objects in the standard way. A line ℓ_0 and an event e_0 not in ℓ_0 determine the plane which is the union of all lines containing e_0 and some point in ℓ_0 . Parallel lines are those that are coplanar and fail to intersect. Finally, linear hypersurfaces are defined as maximal sets of lines intersecting a plane and a given point outside that plane.

With these definitions, the causal construction of the affine structure of Minkowski space is essentially complete. Since the definitions of planes and hyperplanes are standard, they require no additional justifying theorems.

The causal construction of the metric in all its details is a more complicated matter, and is carried out more efficiently within the mathematical

framework of section VI below. Hence the following account is merely a sample of the techniques Robb employed. Nevertheless, the examples chosen (orthogonality and congruence) are of central metrical importance, and the following constructions illuminate their causal-geometrical bases.

Let ℓ_t and ℓ_s be timelike and spacelike lines, respectively, which intersect at an event e_0 . When are we to say that ℓ_t and ℓ_s are orthogonal?

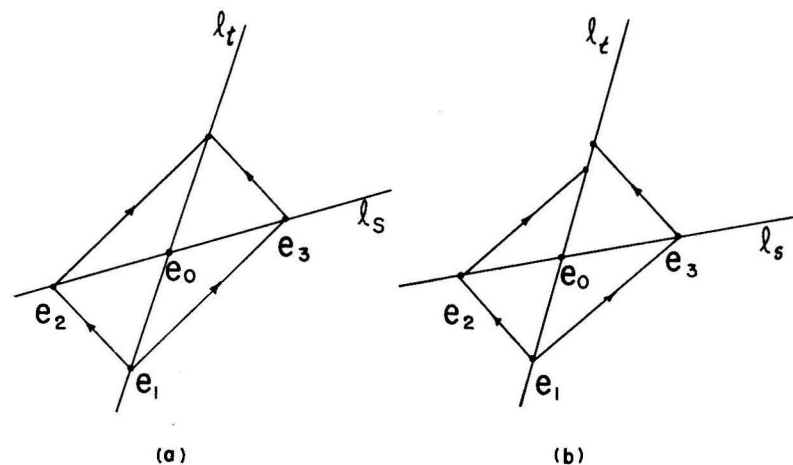


Figure 5.1

Robb's definition is simplicity itself (1921, p. 46ff), and incidentally, helps in reading orthogonality from of a space-time diagram. Choose an event e_1 on line ℓ_t ($e_1 \neq e_0$), as in Figure 5.1(a). The light cone of e_1 will now intersect ℓ_s at events e_2 and e_3 , and each of their cones will in turn intersect ℓ_t once more. If these last intersections are the *same point-event* on ℓ_t (as in (a)), then ℓ_t and ℓ_s are said to be orthogonal.¹⁹ Otherwise (as in (b) below), ℓ_s and ℓ_t are nonorthogonal. If in Figure (a) ℓ_t and ℓ_s are the x^0 and x^1 -axes of an inertial frame, then they do indeed come to "look" orthogonal using this definition. It is not that orthogonal lines in Minkowski space fail to look orthogonal, in contrast to Euclidean lines, but that Minkowskian "orthogonality" bears only a loose family resemblance to its Euclidean counterpart.

Nevertheless, Euclidean spatial orthogonality is now forthcoming. For let ℓ_s^1 and ℓ_s^2 be two spacelike lines which intersect at e_0 . We now determine (define) their orthogonality as follows. Let ℓ_t be the timelike

line through e_0 and orthogonal (in the sense just defined) to both e_1 and e_2 (see Figure 5.2). Then ℓ'_s is orthogonal to ℓ_s^2 just in case every timelike line ℓ'_t in the plane determined by ℓ_t and ℓ_s^2 is orthogonal to ℓ'_s . Remarkably enough, this definition is not only intuitively satisfactory, but can be shown quite rigorously to succeed in reconstructing standard Euclidean orthogonality. (A proof and further discussion are provided in the following section.)

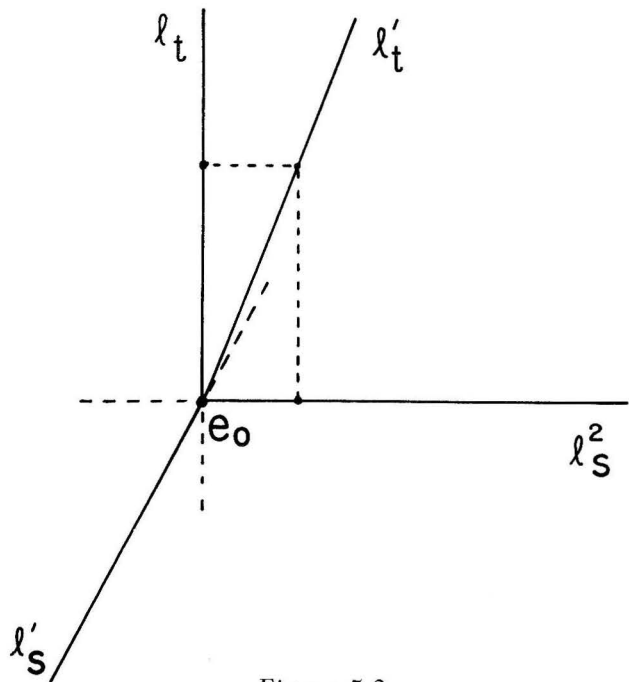


Figure 5.2

The basic idea of this construction (the comparison of a pair of spacelike lines or intervals using a timelike line as an intermediary) is the key to understanding the metrical constructions of causal geometry; the same device is used in the definition of spatial congruence which will soon follow.

Assume now that lines, planes, their parallelism, and orthogonality are given. How does Robb now proceed to construct a theory of congruence from these relations? First of all, consider classical (Euclidean) affine geometry. Within this theory, congruence between segments lying

on *parallel* lines is easily established. Thus if ℓ_1 is parallel to ℓ_2 , AB is a segment on ℓ_1 , and A' is a point on ℓ_2 , then the point B' (on a given side of A') which is such that AB is congruent to $A'B'$ is obtained by projecting parallels AA' and BB' from line ℓ_1 to line ℓ_2 in the usual way. Furthermore, by reflecting parallels from another parallel line (say ℓ_2) the interval AB may be “transported” along its own line as well.

Now just as parallelism allows the definition of a *partial* congruence relation in Euclidean affine geometry, the same is true in space-time geometry. Having developed a theory of parallelism, we define intervals which are the opposite sides of chains of parallelograms to be congruent in space-time.

But what of nonparallel line segments? In Euclidean affine geometry, it is time to add more structure: a congruence predicate and the “compass” axiom, the rotation group, or an inner product, depending upon the general approach adopted. In each case, a new *primitive term and axioms* are added, and in this straightforward sense, the theory of the total congruence relation is *external* to affine geometry (see Klein, 1939, p. 162ff).²⁰

In the causal theory of congruence, there is no need to add new primitives or coordinative decrees, since constructions involving parallelism do not fully exhaust the metrical resources of a causal theory. In addition, there exists a distinguished set of lines, the *causal boundary lines* (the null lines), and these may be used to extend the theory to a full theory of space-time congruence.²¹ Perhaps the best way to see this is to consider some examples.

Let e_1 and e_2 be two events with spacelike separation in two-dimensional Minkowski space-time. Then the light cones of e_1 and e_2 will intersect at two events A and B such that (by Robb’s theory of parallelism) $e_1 B e_2 A$ is a parallelogram. (See Figure 5.3.) Such a parallelogram will be called a “light parallelogram” (Robb’s “optical” parallelograms). Spacelike segments $e_1 e_2$ and timelike segment AB will intersect at some event O , which will be defined as the center of the light parallelogram. The theory of parallelism now allows a proof that $e_1 O$ is congruent to $O e_2$, and OA is congruent to OB . In other words, the diagonals of a light parallelogram bisect each other. So far, then, only parallelism is involved.

But now consider four-dimensional Minkowski space-time (Figure 5.4, drawn with one dimension suppressed). Again let e_1 and e_2 have a spacelike separation. Now there are many light parallelograms with

diagonal e_1e_2 . Choose two with diagonals which are both orthogonal to e_1e_2 , with the result as shown in Figure 5.4. The light parallelograms $e_1 B e_2 A$ and $e_1 B' e_2 A'$ share the diagonal e_1e_2 and center O . The remaining half-diagonals OB and OB' are now *defined* as congruent. This idea may now be extended to provide a general definition of congruence of timelike intervals along the following lines.

Suppose we are given intersecting timelike lines ℓ_1 and ℓ_2 , with segment OB on ℓ_1 , and we wish to locate B' , i.e., the point on the "B-side" of ℓ_2 such that OB' is congruent to OB . The line with segment e_1e_2 (see Figure 5.4) can be shown to be uniquely determined by the condition that it be orthogonal to both ℓ_1 and ℓ_2 . Point-events e_1 and e_2 are now determined by the "backwards" light cone of event B . The "forward" cones of e_1 and e_2 (actually only one is necessary) now determine point B' on ℓ_2 ; in effect, the "top half" of the intersecting light parallelograms is determined.

With this construction, the problem of the congruence of arbitrary

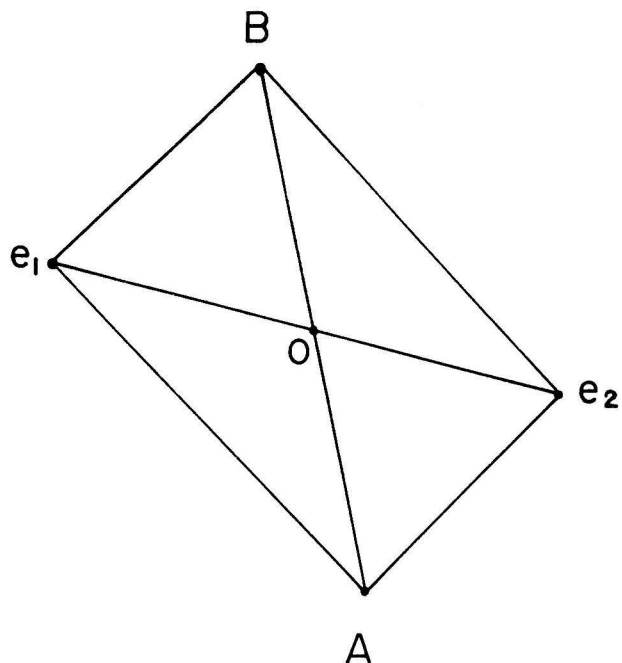


Figure 5.3

timelike intervals is all but solved. Timelike intervals AB and CD on *intersecting* lines are compared by using the above construction to transfer segment AB from its line to the other. Parallelism now allows the transported segment to be compared with segment CD . If intervals AB and CD lie on *nonintersecting* lines, we merely use a third line intersecting the second (CD line) and parallel to the first (AB line) as an intermediary and proceed as before.

An attractive feature of this construction is that it reveals the causal basis of the hyperbolas which form the contour lines of the Lorentz metric. Thus in Figure 5.5(a), let OA be the unit timelike interval. Then the locus of all events later than O and also a unit distance from O is given by the familiar hyperbola shown in the figure. The causal theory of congruence provides some geometrical insight into the situation. Thus let OA be

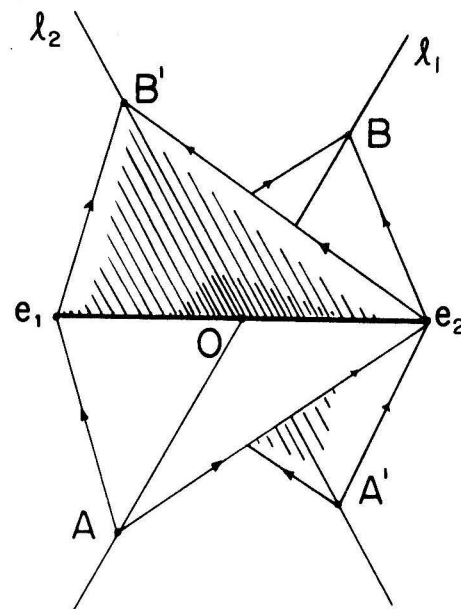


Figure 5.4

a unit interval and consider the problem of finding A' on an arbitrary timelike line ℓ through O and in the $t-x$ plane (see Figure 5.5(b)). By construction, the y -axis is orthogonal to both the t -axis and line ℓ . The light cone of A now determines e_1 and e_2 , and so the diagonal of the top half of a light parallelogram. But the e_1 and e_2 cones will both intersect line ℓ in a unique point A' , which, by our earlier definition, yields OA' congruent to OA .

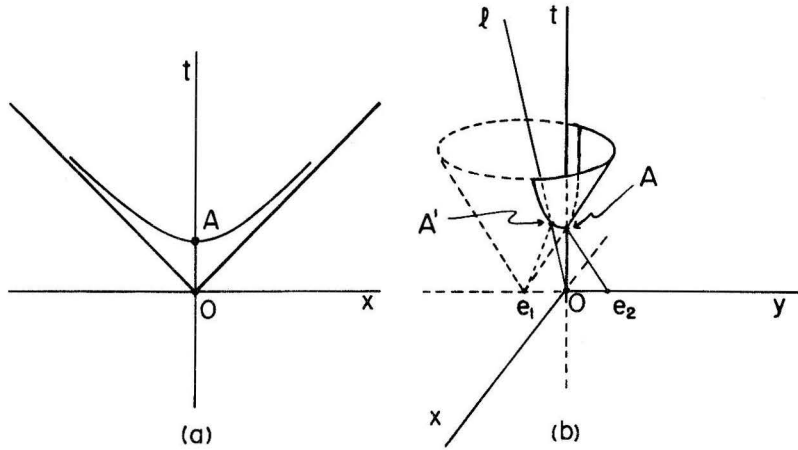


Figure 5.5

Now consider *all* timelike lines like ℓ that intersect O and lie in the $t-x$ plane. Use only the light cone of e_1 to determine the unit distance on this line, since the e_2 cone gives the same result. Then the intersection of the e_1 cone with such lines is the intersection of the e_1 cone with the $x^0 - x^1$ plane, which is just the Minkowski hyperbola.²²

Finally, let us consider the congruence of spacelike intervals. These are compared by first associating them with a timelike interval, and comparing the latter as above. Once again, the general idea is to use light parallelograms, and then compare diagonals. For example, let AB and $A'B'$ below be spacelike intervals (see Figure 5.6). Let $ADBC$ be a light parallelogram with AB as diagonal (there will be many such; pick any one), and let $A'D'B'C'$ be chosen similarly. Then diagonals DC and $D'C'$ will be timelike (theorem), and AB is congruent to $A'B'$ if and only if (definition) DC is congruent to $D'C'$. The unique congruences thus specified probably

yield a Euclidean geometry for both spacelike planes and linear three-dimensional spacelike subspaces of space-time (Robb, 1921, p. 67ff).

Of course, this last group of constructions requires supporting proofs of accuracy; these are supplied by Robb (1914). Nevertheless, with these constructions in mind we may stand back and ask: how do they manage to succeed?

Perhaps the two most remarkable features of the causal geometry of flat space-time are: first, the ability of the causal relations to yield the affine structure of space-time; and second, the lack of necessity for additional primitives (or coordinative definitions) in order to obtain the congruences of the Lorentz metric. From the Euclidean, or even Riemannian, viewpoint, this feat appears almost magical. How does causal geometry do the trick?

To answer this question we must go back to the conical structure of the causal space of each event. The set of events causally connectible with a given event e (the causal space of e) has a boundary or shell, the causal cone of e . The causal cone of e is definable in terms of $<$ alone, and light lines (better called "causal boundary lines" since we need not suppose them to be occupied) are then defined as above. The light lines form the foundation for the linear structure of Minkowski space-time.

Similarly, the *light* parallelograms form the basis for the comparison of nonparallel segments. A comparison with Euclidean space may be useful. Let AB be a segment in Euclidean three-space. Then the Euclidean parallelograms which have AB as their common diagonal form a motley

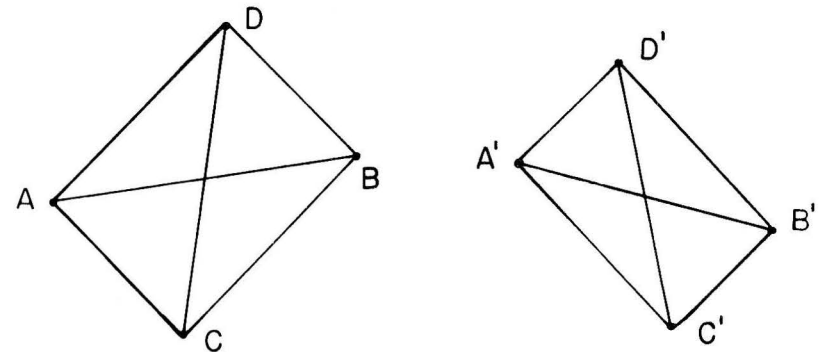


Figure 5.6

set whose non- AB diagonals have little in common aside from the condition just used to define them. But the *light* parallelograms of Minkowski space with a given diagonal AB form a severely restricted set. For one thing, they all have a common center (1914, theorem 59; 1921, p. 31). Hence the initial *plausibility* of defining corresponding half-diagonals of such light parallelograms as congruent. What is remarkable is that this definition is compatible with the affine congruence relations so that both taken together yield a congruence relation which is provably an equivalence relation on its domain.

Next consider the question of spatial geometry from the standpoint of the causal theory. As we have just seen, the congruence of spacelike intervals is defined uniquely in the causal theory by comparison with timelike intervals. The spatial geometry of a spacelike plane (or spacelike, three-dimensional, linear subspace of space-time) is then determined by the space-time congruences and linear structure of the events in the plane (subspace). The resulting geometry is then provably Euclidean.

At this point, an interesting issue arises. As Grünbaum has pointed out, although a metric yields a unique geometry on a space, the converse is, in general, false. In particular, the same set of points may be provided with two Euclidean metrics which nevertheless differ in their congruence verdicts (1973, p. 98 ff.).²³ However, the causal constructions just outlined yield a Euclidean spatial geometry by *first* delivering the congruence verdicts of that geometry! Thus, although the existence of alternative Euclidean metrics is a mathematical possibility, it is not, within special relativity, a physical possibility!

There is, of course, an important gap in this argument. For it has only been shown (or more precisely, claimed) that *given* the space-time geometry, spatial congruence is uniquely determined. Does Minkowski geometry itself allow alternative congruences? If so, we might expect this fact to reinstate derivatively alternative spatial congruences as well.

Clearly the preceding constructions show that this situation does not arise. For space-time congruence is *explicitly definable* in terms of causal structure. Given that this (causal) structure is fixed, space-time congruence—and hence spatial congruence—are uniquely determined by the preceding constructions.²⁴ Thus *the causal structure of Minkowski space-time allows the unique determination of spatial congruence within its spatial linear hyperspaces*. As a result, no extra-causal resources such

as extended rods are needed in order to specify spatial congruence. On the contrary, the causal definition of spatial congruence may be used to provide an ontological basis for the assessment of the rigidity or accuracy of material congruence standards.

The preceding constructions were not, in general, accompanied by justifying theorems. However, the development of the linear theory was reasonably rigorous. The metrical constructions (orthogonality and congruence) were not provided with proofs of adequacy, and were, in any case, incomplete. The following section repairs this defect by defining the Minkowski metric in terms of its linear and causal structure, with proofs of adequacy provided. Fortunately, this development may be carried out in a reasonably direct way within the framework of modern metric affine geometry.

VI. From Affine Structure to the Minkowski Metric

In the preceding section it was shown that the construction of a space-time metric falls naturally into two parts: (1) the theory of linear or affine structure; and (2) the theory of congruence and metric. Constructions that yield the affine structure of Minkowski space-time have been provided and justified. However, the causal construction of the Minkowski metric was merely outlined. In this section, the construction of the Minkowski metric from parallelism will be developed in detail along lines proposed by Robb, but using the techniques of the modern theory of metric affine spaces. The aim is to show that, given the usual linear structure of affine spaces, the mere addition of the Minkowski null cone makes possible the construction of the metric structure of Minkowski space-time: orthogonality, congruence, and interval. For Minkowski space-time, the transition from parallelism to metric—which, in Euclidean spaces, is carried out by adding congruence postulates or an inner product—may be achieved merely by adding the null cone structure. We begin by outlining the general structure of affine metric spaces.²⁵

Let V be a vector space over the field of real numbers. A (nonsingular) *metric vector space* is obtained by providing V with an *inner* product (\langle, \rangle) which is just a function that assigns a real number to every pair of vectors in V , and is, in addition, linear, symmetric, and nonsingular.²⁶ Thus a metric vector space²⁷ may be regarded as a pair $V_I = (V, \langle, \rangle_I)$ with V a real vector space, and \langle, \rangle_I an inner product on V .

The choice of this inner product now determines the geometry of V_I . For example, if we now require that the inner product be *positive definite*, i.e.,

$$\text{for all } v \in V, \langle v, v \rangle \geq 0, \text{ with} \\ \langle v, v \rangle = 0 \text{ iff } v = 0$$

the resulting metric vector space is said to be Euclidean.²⁸ *Minkowski vector space* (V_m^4) may be defined as a four-dimensional (dimension here is that of the vector space V) real metric vector space whose inner product satisfies the following condition:

(M) There are four linearly independent vectors

$$v_0, \dots, v_3 \text{ such that} \\ \langle v_0, v_0 \rangle < 0, \text{ and } \langle v_k, v_k \rangle > 0, k = 1, 2, 3.²⁹$$

The *light cone* or *null cone* of V_m^4 consists of all vectors v , called *null vectors*, whose length² ($\langle v, v \rangle$) is zero. Clearly the zero vector, $\vec{0}$, is in the null cone of V_m^4 , and it is easy to show that the null cone consists of a three-dimensional surface (not subspace, however) in V_m . A vector v is said to be *timelike* if $\langle v, v \rangle < 0$ and *spacelike* if $\langle v, v \rangle > 0$. Two vectors u and v are *congruent* if and only if they have the same length, i.e., $\langle u, u \rangle = \langle v, v \rangle$. Notice that only timelike vectors are congruent to timelike vectors, and so too for spacelike vectors.³⁰ All null vectors are hereby made congruent, but this is merely a convenience. Null vectors may be compared *along a given null line* by means of the scalars of V^4 . (Thus av and $-av$ might be held congruent, av and $2av$ not congruent, even when v is null.)

The above account outlines the standard approach to Minkowski metric vector spaces. A specific inner product is added to V^4 , and then used to define vector length², the light cone, spacelike and timelike vectors, and, if we like, congruence of vectors. The analogue of Robb's approach is quite different. Here we begin with the vector space V^4 and its light cone, which will be called "Null." Congruence, orthogonality, and finally, an inner product are now defined in terms of this structure. In other words, where the standard account begins with a structure of type (V^4, \langle, \rangle) , and goes on to define Null, a subset of V^4 , as:

$$(1) \text{ Null} \equiv \{v \mid v \in V^4 \text{ and } \langle v, v \rangle = 0\},$$

the following account defines congruence and the inner product function in terms of Null (and the vector space structure of V^4).³¹ However, before

proceeding, a few remarks are in order concerning the relationship of the above Minkowski metric *vector space* to affine Minkowski space-time, the universe of the latter consisting of points (or point-events), that of the former being a set of vectors.

Let X be a nonempty set of points. Recall that $V_I = (V, \langle, \rangle_I)$ is a metric *vector space*. We now make the set of points X into a geometrical space (*a real affine metric space*) by using the metric *vector space* V_I to induce a metric on X . The obvious way to do this is to associate each pair of points in X with a vector in V . Thus if $\langle x, y \rangle$ is an ordered pair of points in X , the vector in V associated with $\langle x, y \rangle$ is designated by " $\langle \vec{x}, \vec{y} \rangle$." The distance² between x and y in X is now defined in a natural way as the length² of $\langle \vec{x}, \vec{y} \rangle$ in V_I (i.e., as $\langle \langle \vec{x}, \vec{y} \rangle, \langle \vec{x}, \vec{y} \rangle \rangle$). The straight line through x and y is given by the set of points associated with all vectors $a \langle \vec{x}, \vec{y} \rangle \in V, a \in \text{Reals}$, etc.

Now the association of point-pairs in X with vectors requires an *association function* A , usually called the *action* (function) of V_I on X . Of course, A must induce the structure of V_I on X in a "natural" way, and is accordingly subjected to the following requirements:

- (2) A is a function from $V \times X$ to X , such that:
 - (a) $A(\vec{0}, x) = x$, for all $x \in X$,
 - (b) If $A(v, x) = A(w, x)$, then $v = w$, and
 - (c) $A(u + v, x) = A(u, A(v, x))$. (Cf. Snapper and Troyer, 1971, p. 6.)

A *real affine metric space* is now defined as a set X , and a real metric vector space V_I together with an action function A .

Now consider two such affine spaces, (X, V_I, A) and (Y, V'_I, A') . It is not difficult to show that these spaces are isometric if and only if V_I and V'_I are isometric (Snapper and Troyer, 1971, p. 387, ex. 12). This amounts to saying that the study of an affine metric geometry essentially reduces to the study of its metric *vector space* V_I . In short, while an affine space (say Minkowski space-time) and its associated metric vector space (say V_m^4) are not identical, the geometric structure of the affine space derives entirely from its associated metric vector space, so the geometry of the former essentially reduces to that of the latter. Hence Minkowski space-time, i.e., affine Minkowski metric space, is defined as a triple (M, MV^4, A) , where M is the set of its "points" (or point-events), $MV^4 = (V^4, \langle, \rangle_m)$, and A specifies the action of MV^4 on M in accord with the requirements of (2) above.

The above isomorphism theorem shows that, having distinguished Minkowski space-time from its metric vector space, we now may proceed to concentrate on the latter. The first task will be to use the Null vectors of V^4 to define congruence of its vectors, and then to go on to show how to define orthogonality and a Minkowski inner product on V^4 .

The central notion in the following is that of the *conjugate class*, $\text{Conj}(v)$, of a vector v . A vector u is in the conjugate class of v if and only if both u and $-u$ yield a null vector when added to v . It turns out (proposition 6.1 below) that the conjugate class of v contains just those vectors that are orthogonal to v and have length² equal to negative length² of v .

Definition 6.0. Where $v \in V^4$,

$\text{Conj}(v) = \text{df. } \{u \mid u \in V^4 \text{ and } v + u, v - u \in \text{Null}\}.$

Proposition 6.1. For all $u, v \in V^4$,

$u \in \text{Conj}(v)$ iff $\langle u, v \rangle = 0$ and $\langle u, u \rangle = -\langle v, v \rangle.$

Proof: (a) Suppose $u \in \text{Conj}(v)$. Then (1) $\langle u + v, u + v \rangle = 0$

and (2) $\langle u - v, u - v \rangle = 0$. Expanded, (1) and (2) become:

(3) $\langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle = 0$ and (4) $\langle u, u \rangle - 2\langle u, v \rangle + \langle v, v \rangle = 0.$

Adding and subtracting (3) and (4) yields:

$\langle u, u \rangle = -\langle v, v \rangle$ and $\langle u, v \rangle = 0.$

(b) As in case (a). Done.

The following two propositions will be used later in proving the accuracy of the forthcoming congruence definition:

Proposition 6.2. (a) $\text{Conj}(\vec{0}) = \text{Null}.$

(b) If $u \in \text{Null}$ and $u \neq \vec{0}$, then $\text{Conj}(u) = \langle\langle u \rangle\rangle$, the subspace of V^4 spanned by u .

Proof: (a) Immediately from definition 6.0.

(b) Suppose $v \in \text{Conj}(u)$. Then by proposition 6.1, $v \in \text{Conj}(u)$ iff (1) $\langle v, v \rangle = -\langle u, u \rangle = 0$ and (2) $\langle u, v \rangle = 0$. Hence if $v = \vec{0}$, $v \in \text{Conj}(u)$, and thus in $\langle\langle u \rangle\rangle$. Suppose then that $v \in \text{Conj}(u)$, $v \neq \vec{0}$ and $v \notin \langle\langle u \rangle\rangle$. Then u and v are linearly independent, and so $\langle\langle u; v \rangle\rangle$, the subspace spanned by u and v , is a plane in V_m^4 . Let w be in $\langle\langle u; v \rangle\rangle$. Then $w = au + bv$, for some $a, b \in \text{Reals}$. However, by virtue of (2), $\langle w, w \rangle = 0$. Hence $\langle\langle u; v \rangle\rangle$ is a null plane in V_m^4 , which contains no such planes (see Snapper and Troyer, 1971, corollary 195.2). Thus we must have $v \in \langle\langle u \rangle\rangle$.

On the other hand, if $v \in \langle\langle u \rangle\rangle$, $v = au$, so $\langle v, v \rangle = a^2 \langle u, u \rangle =$

0 , and $\langle u, v \rangle = a \langle u, u \rangle = 0$. Hence by proposition 6.1, $v \in \text{Conj}(u)$. Done.

Next it is shown that a vector and its negative have the same conjugates, and that every vector has some conjugate.

Proposition 6.3. (a) For all u , $\text{Conj}(u) \neq \{ \}$

(b) For all u , $\text{Conj}(u) = \text{Conj}(-u)$.

Proof: (a) If $u \in \text{Null}$, then the result follows from proposition 6.2. Suppose then that u is timelike. Let v be spacelike and hence linearly independent of u . Then $\langle\langle u, v \rangle\rangle$ is a Lorentz plane and contains a spacelike vector w orthogonal to u . Multiply w by the appropriate scalar a to obtain aw : $\langle aw, aw \rangle = -\langle u, u \rangle$, and $\langle u, aw \rangle = 0$. Proposition 6.1 now yields $aw \in \text{Conj}(u)$. If we suppose u to be spacelike, we choose a timelike vector v' and proceed similarly.

(b) Immediately from definition 6.0. Done.

We are now in a position to define congruence of vectors in V^4 . Note that by proposition 6.1, conjugate vectors are not in general congruent (their lengths being of opposite sign), but for any vector u , the vectors in $\text{Conj}(u)$ are all congruent to each other. Still, if u and v are, say, congruent timelike vectors, can we count on there being a spacelike vector w such that $u \in \text{Conj}(w)$ and $v \in \text{Conj}(w)$, thereby yielding their congruence? The following definition relies on this being the case, and the propositions which follow justify this confidence.

Definition 6.4. $u \simeq v$ (u is congruent to v) = df.

$\text{Conj}(u) \cap \text{Conj}(v) \neq \{ \}.$

Proposition 6.5. For all u, v ,

(1) $u \simeq u$ (reflexivity),

(2) if $u \simeq v$, then $v \simeq u$ (symmetry).

Proof: (1) Immediately from definition 6.4 and proposition 6.3.(a).

(2) Immediately from definition 6.4.

The following proposition now establishes the accuracy of the congruence definition.

Proposition 6.6. $u \simeq v$ iff $\langle u, u \rangle = \langle v, v \rangle.$

Proof: (a) Assume $u \simeq v$. Then for some vector w_0 , $w_0 \in \text{Conj}(u)$ and $w_0 \in \text{Conj}(v)$. Hence by proposition 6.1, $\langle u, u \rangle = -\langle w_0, w_0 \rangle = \langle v, v \rangle.$

(b) Suppose $\langle u, u \rangle = \langle v, v \rangle.$

John A. Winnie

- (1) $u, v \in \text{Null}$. By proposition 6.2, $\vec{0} \in \text{Conj}(u)$ and $0 \in \text{Conj}(v)$, whether or not u and v are the zero vector.
- (2) $u, v \notin \text{Null}$.
- (α) u, v linearly dependent. Then $u = kv$, and thus $\langle u, u \rangle = k^2 \langle v, v \rangle$. But $\langle u, u \rangle = \langle v, v \rangle$, so $k = \pm 1$.

Hence $u = v$ or $u = -v$. In either case, by proposition 6.3, $\text{Conj}(u) \cap \text{Conj}(v) \neq \{ \}$.

- (β) u, v linearly independent. Since $\langle u, u \rangle = \langle v, v \rangle$, and $u, v \notin \text{Null}$, u and v are either both timelike or both spacelike. Hence $\langle\langle u, v \rangle\rangle$ is either a Lorentz plane or a Euclidean plane. In either case $V^4 = \langle\langle u, v \rangle\rangle \oplus \langle\langle u, v \rangle\rangle^*$, where $\langle\langle u, v \rangle\rangle^*$ is the orthogonal complement of $\langle\langle u, v \rangle\rangle$. (See Snapper and Troyer, 1971, p. 155, ex. 4b.) Thus $\langle\langle u, v \rangle\rangle^*$ is either a Euclidean plane or a Lorentz plane, respectively. Hence when u, v are both timelike, a spacelike vector w_0 is in $\langle\langle u, v \rangle\rangle^*$ and is thus orthogonal to both u and v . Multiplication by an appropriate scalar will now yield an aw_0 , $\langle aw_0, aw_0 \rangle = -\langle u, u \rangle = -\langle v, v \rangle$. Hence by proposition 6.1, $aw_0 \in \text{Conj}(u) \cap \text{Conj}(v)$. When u, v are both spacelike, a similar argument succeeds. Done.

Now that it has been shown that congruence of vectors is definable using the null cone of V_m^4 , the orthogonality of vectors in V_m^4 is defined next. We want our definition to yield u orthogonal to v just when $\langle u, v \rangle = 0$. With congruence and orthogonality defined, the construction of the Minkowski inner product on V^4 is simple and direct.

As an aid to the constructions to follow we now define what it means for two vectors u and v to be of the same type ($u \longleftrightarrow v$)—intuitively, to be either both null, both timelike, or both spacelike. The definition is obvious. The proposition that follows provides justification.

Definition 6.7. $u \longleftrightarrow v = \text{df.}$ for some $a \neq 0$, $au \simeq v$.

Proposition 6.8. $u \longleftrightarrow v$ iff $u, v \in \text{Null}$, or u, v timelike, or u, v spacelike.

Proof: Immediately from proposition 6.6 and from the fact that if u and v are of the same type, a suitable choice of $a \neq 0$ will yield $\langle au, au \rangle = \langle v, v \rangle$.

By the *type of a vector* v we shall mean the set of all vectors of the same type as v . However, the above definition has the zero vector different in type from any spacelike or timelike vector; so for the sake of convenience we add it to each class.

Definition 6.9. Where v is any vector, the type of v (Type(v)) is just the set of all vectors of the same type as v , together with the zero vector, i.e.,

$$\text{Type}(v) = \text{df.} \{u \mid u \longleftrightarrow v \text{ or } u = \vec{0}\}.$$

The construction of orthogonality is helped by being able to distinguish timelike and spacelike vectors. It is well known (cf. Noll, 1964) that the set of spacelike vectors contains a subspace of dimension three, whereas the timelike vectors contain subspaces of dimension one at most. We use this fact below to distinguish the two sorts of vectors.

Definition 6.10. Where u is any vector,

- (a) u is timelike = df. Type(u) contains subspaces of at most dimension one;
- (b) u is spacelike = df. Type(u) contains subspaces of dimension three.

We are now in a position to define orthogonality of vectors in v .

The definition of orthogonality is made easier by recalling that in Minkowski spaces: (1) timelike vectors are only orthogonal to spacelike vectors, (2) two (nonzero) null vectors are orthogonal if and only if they are linearly dependent, and (3) the zero vector is orthogonal to every vector (cf. Trautman, 1965, section 3.2). Hence only three interesting cases of orthogonality remain: spacelike vectors orthogonal to (nonzero) null vectors, timelike vectors, and other spacelike vectors. The case of the orthogonality of a spacelike and a timelike vector amounts to a simple extension of the idea of conjugate vectors introduced above. The geometrical significance of the other two cases of orthogonality is quite different, however, and each case merits individual treatment. First, the case of two orthogonal spacelike vectors.

The intuitive idea is this. Let u and v be orthogonal spacelike vectors. Then there will be a plane $\langle\langle w_0, v \rangle\rangle$, containing v , all of whose vectors (and hence all of whose timelike vectors) are orthogonal to u (see Figure 5.2).

The orthogonality of a (nonzero) null vector and a spacelike vector has still another significance. Let u and v be the null and spacelike vector, respectively. Then u and v may be such that their "tips" are not causally connectible, i.e., $u - v$ is spacelike. If this is so *no matter how* v is extended (for all av , $a \neq 0$), then and only then are u and v orthogonal. In other words, an intersecting spacelike and a null line are orthogonal just in case no point on the spacelike line, save the intersection point itself, is

causally connectible with any point on the null line (cf. Robb, 1921, p. 46, definition B).

These intuitive explanations are stated more precisely in the following definition. The succeeding proposition shows that the construction is successful.

Definition 6.11. $u \perp v = \text{df.}$

- (1) $u = 0$ or $v = 0$, or
- (2) $u, v \in \text{Null}$, $u \neq 0 \neq v$, and $u = av$, for some $a \in \text{Reals}$, or
- (3) u is spacelike and:
 - (a) v is timelike and for some $a \in \text{Reals}$, $a \neq 0$, $au \in \text{Conj}(v)$; or
 - (b) v is spacelike and for some $w_0 \in \text{Conj}(v)$, if t is timelike and in $\langle\langle w_0; v \rangle\rangle$, then $u \perp t$, i.e., t satisfies (a); or
 - (c) v is null, $v \neq 0$, and $v - au$ is spacelike for all $a \in R$, $a \neq 0$; or
- (4) as in (3), with u and v interchanged.

Proposition 6.12. $u \perp v$ iff $\langle u, v \rangle = 0$.

Proof: (A) Suppose $\langle u, v \rangle = 0$. By the remarks above, definition 6.11 covers all possible cases. Hence we need only show its accuracy in each case. Cases (1) and (2) are trivial.

Case (3) (a): Choose $a \neq 0$ such that $\langle au, au \rangle = -\langle v, v \rangle$.

Then by proposition 6.1, $au \in \text{Conj}(v)$.

Case (3) (b): Since $\langle u, v \rangle = 0$, u and v are linearly independent, so $\langle\langle u; v \rangle\rangle$ is a Euclidean plane. Let w_0 be a timelike vector, $\langle u, w_0 \rangle = 0$, $\langle v, w_0 \rangle = 0$, suitably chosen so that $\langle w_0, w_0 \rangle = -\langle v, v \rangle$. By proposition 6.1, $w_0 \in \text{Conj}(v)$. Now suppose t timelike and $t \in \langle\langle w_0; v \rangle\rangle$. Then $t = aw_0 + bv$.

Hence $\langle u, t \rangle = a\langle u, w_0 \rangle + b\langle u, v \rangle = 0$. Choose $d \neq 0$, so that $\langle dt, dt \rangle = -\langle u, u \rangle$. Then $\langle dt, u \rangle = 0$ and by proposition 6.1, $dt \in \text{Conj}(u)$. Case (3) (c): By a standard theorem, (i) $\langle v - au, v - au \rangle = \langle v, v \rangle - 2\langle v, au \rangle + a^2\langle u, u \rangle$.

Since $\langle v, v \rangle = \langle v, u \rangle = 0$, (i) becomes (ii) $\langle v - au, v - au \rangle = a^2\langle u, u \rangle$.

Since $\langle u, u \rangle > 0$, (ii) shows that $v - au$ is spacelike for all $a \neq 0$. Furthermore, (i) shows that unless $\langle v, u \rangle = 0$, some choice of $a \neq 0$ will make $\langle v - au, v - au \rangle < 0$, proving sufficiency as well.

Case (4): as in case (3).

(B) Suppose $u \perp v$. Cases (1) and (3) are trivial. Case (3) (a) follows from proposition 6.1. Case (3) (c) was shown under (A) above. Case (4) is treated like case (3). This leaves only Case (3) (b). Choose $t_1, t_2 \in \langle\langle w_0;$

$v \rangle\rangle$ (a Lorentz plane) to be timelike and linearly independent. Then by (3) (b), $u \perp t_1$, $u \perp t_2$, so by case (3) (a), $\langle u, t_1 \rangle = \langle u, t_2 \rangle = 0$. Now $v = at_1 + bt_2$ (since t_1, t_2 a basis of $\langle\langle w_0; v \rangle\rangle$), so $\langle u, v \rangle = a\langle u, t_1 \rangle + b\langle u, t_2 \rangle = 0$. Done.

A Minkowski inner product is now defined as follows. First, choose a timelike vector t_0 as a unit. Let s_1 be any vector in $\text{Conj}(t_0)$. Hence $t \perp s_1$ and $\langle s_1, s_1 \rangle = -\langle t_0, t_0 \rangle$. We now complete the set by adding s_2, s_3 , both spacelike, congruent to s_1 , and such that the vectors $B = \{t_0, s_1, s_2, s_3\}$ are pairwise orthogonal. (Since this is V_m^4 , we know that such exist.) The set B is thus a Lorentz basis of V_m^4 , so we may define the inner product \langle, \rangle_m^* in the usual way:

Definition 6.13. $\langle u, v \rangle_m^* \equiv -u^0v^0 + u^k v^k$, $k = 1, 2, 3$, where u^α and v^α are the components of u and v in any basis $B = \{t_0, s_1, s_2, s_3\}$, t_0 is timelike, $s_1 \in \text{Conj}(t_0)$, $s_3 \simeq s_2 \simeq s_1$, and the members of B are pairwise orthogonal.

The following theorem follows automatically:

Proposition 6.14. Let \langle, \rangle^* be a Minkowski inner product on V^4 , with null cone = Null and $\langle t_0, t_0 \rangle^* = -1$. Then $\langle u, v \rangle^* = \langle u, v \rangle_m^*$.

Proof: Since the null cone of \langle, \rangle^* is Null, all previous theorems apply.

Let $B = \{t_0, s_1, s_2, s_3\}$ be as specified in definition 6.13. Its members are mutually orthogonal \perp , and so, by theorem 6.12, have pairwise inner products (\langle, \rangle^*) equal to 0. But $\langle t_0, t_0 \rangle^* = -1$, and so by propositions 6.1 and 6.6, $\langle s_k, s_k \rangle^* = +1$, $k = 1, 2, 3$. Thus B is a Lorentz frame and $\langle u, v \rangle^* = -u^0v^0 + u^k v^k \equiv \langle u, v \rangle_m^*$. Done.

So, given the vector space V^4 and the Minkowski null cone in V^4 , the Minkowski metric may be uniquely defined (up to a unit of length). In Euclidean geometry, on the other hand, the choice of a unit vector does not suffice for the construction of the metric. In addition, a set of $n - 1$ linearly independent vectors must be exhibited and defined to be mutually congruent and pairwise orthogonal. The arbitrariness of the resulting metric, from the physical standpoint, will reside in the extent to which the choice of these other vectors is based upon factual physical relations. When Minkowski geometry is seen in a similar way, the choice of s_1, s_2 , and s_3 comes under scrutiny, and issues involving the conventionality of the spatial length of transported spatial measuring standards will naturally arise. The above construction shows how this difficulty is eliminated

within Minkowski space-time. In addition to the vector t_0 , we obtain the metric by exhibition, but now by the exhibition of the null cone in V^4 . However, the null cone, now interpreted as the boundary of the causal connectedness relation, is free from the conventions that attach to the choice of various spatial and temporal measuring devices.

The last theorem (proposition 6.14) is also a step on the way toward fulfilling the claim made earlier (section V) that two Minkowski metrics on the same set M which have identical null cones are identical up to a constant scale factor. However, proposition 6.14 assumes that the two metrics yield the same linear structure for space-time, and this is an unnecessary restriction, since two Minkowski metrics with the same null geodesics agree on all geodesics. Hence the following theorem is just what is needed:

Proposition 6.15. Let $\mathfrak{M}_1 = \langle M, \eta \rangle$ and $\mathfrak{M}_2 = \langle M, \eta' \rangle$ be Minkowski space-times³² with identical null cone structures. Then $\eta' = k\eta$, where k is a positive constant.

Proof: Since \mathfrak{M}_1 and \mathfrak{M}_2 have the same null cone structure, they are conformal, i.e., $\eta' = \sigma\eta$, where σ is a positive real-valued, smooth function defined on all of M . Since \mathfrak{M}_1 and \mathfrak{M}_2 are both flat, it can be shown—as in Haantjes (1937), p. 702—that σ must satisfy

(1) $2S_{\lambda, \mu} = S_{\mu}S_{\lambda} - (\eta_{\mu\lambda}/2)(\eta^{\alpha\beta}S_{\alpha}S_{\beta})$, where $S_i \equiv \partial \ln \sigma / \partial x^i$. Equations (1) may be solved, giving $\sigma = \text{constant}$ as a solution, along with solutions of two other types. However, for each of these types, σ is not defined on *all* of M (Haantjes, 1937, section 2). Hence $\sigma = \text{constant}$. Done.

It now follows that the causal structure of Minkowski space-time fixes its metric up to a constant factor. As a result the spatial metric of the inertial frames of special relativity is determined (up to a constant factor) as well.

VII. The Causal Definition of Congruence

While the constructions of synthetic causal geometry move deeply into the structure of Minkowski space, the homogeneous geometry of Minkowski space-time permits the effective use of group-theoretic methods in its study. These methods provide a considerably easier way to show the causal definability of congruence in Minkowski space. For E. C. Zeeman has shown (1964) that the causal group of flat space-time and its "Lorentz" group are one and the same; in this section it will be shown that the causal

definability of Minkowski space-time congruence is a relatively easy consequence of Zeeman's result. To begin with, let us be quite clear about what Zeeman's theorem asserts.

Recall that " $TB(e_1e_2e_3)$ " asserts that event e_3 is *chronologically between* events e_1 and e_3 , and " $CB(e_1e_2e_3)$ " asserts that e_2 is *causally between* events e_1 and e_3 . Now a *symmetry* (or automorphism) of a given structure, say $\langle M, TB \rangle$ or $\langle M, CB \rangle$, is a bijective (one-one onto) mapping of its universe M to itself which, in addition, preserves certain (perhaps all) of that structure's relations. The set of all such symmetries is easily shown to be a group when functional composition is taken as the group operation, and so is called a "symmetry group" of that structure. The symmetry groups of $\langle M, TB \rangle$ and $\langle M, CB \rangle$ will henceforth be called the *chronology group* (TG) and the *causality group* (CG) of Minkowski space-time, respectively. Thus TG consists of all one-one mappings of Minkowski space-time onto itself that preserve chronological betweenness, while CG contains just those mappings that preserve causal betweenness.³³

However, in Minkowski space-time, the chronology relation \ll and the causality relation $<$ are interdefinable (see Kronheimer, 1967, section 2). It follows immediately that a mapping which preserves the one will preserve the other, and so the chronology group and causality group are identical ($CG = TG$).

Of course, this easy result is not Zeeman's theorem. What Zeeman has shown is that the causality-chronology group is in turn identical with the set of all mappings that preserve the Minkowski metric up to a scale factor. This group thus contains the metrical symmetries together with those maps that uniformly expand or contract Minkowski space-time ("dilations," "magnifications"). For the sake of clarity, let us define the groups involved more carefully.

Let (M, I^2) be the points of Minkowski space-time together with the interval function $I^2: M \times M \rightarrow R$, where $I^2(e_1, e_2)$ is the interval squared between e_1 and e_2 . Then an *isometry* of M onto M is a one-one mapping of M onto M which preserves I^2 . The set of all such isometries forms a symmetry group, called the Poincaré Group (PG) of M .³⁴

It would be too much to expect the causal group CG to be identical with the Poincaré group of flat space-time, since the Poincaré group fixes the scale of space-time geometry, and this would seem to require extra-causal considerations or conventions. In any case, the geometrically fundamental group is not the isometry group, but rather the *similarity group*, of a

geometry: the set of all one-one mappings of the space onto itself that preserve the metric up to a scale factor.³⁵

Thus a mapping h is a *similarity* of Minkowski space if and only if h is a bijection of M to M such that:

$$I^2(e_1, e_2) = j \cdot I^2(h(e_1), h(e_2)),$$

for some positive real number j . The group of all similarities of Minkowski space includes the Poincaré Group as a subgroup (take $j = 1$), and will henceforth be called the Extended Poincaré Group (*EPG*). A theorem on metric affine spaces (Snapper and Troyer, 1971, theorem 416.1) now tells us that every similarity is the product of a magnification and an isometry, so we may characterize *EPG* as obtained from the Poincaré group by adding magnifications and closing under functional composition.³⁶

We are now in a position to state Zeeman's theorem:

(Z) The causality (= chronology) group of Minkowski space is just the Extended Poincaré Group; i.e., $CG = TG = EPG$.

In other words, any one-one mapping of Minkowski space-time onto itself that preserves its causal structure also preserves its metric up to a scale factor. (The converse is obvious since $<$ and $<<$ are definable in terms of I^2 .)

Clearly Zeeman's theorem shows that the metrical and causal structure of space-time are very closely related. According to the theorem, the similarity group of the structure $\langle M, I^2 \rangle$ and the symmetry group of $\langle M, CB \rangle$ are identical. However, exactly what does this tell us about the connection between the relation CB of causal betweenness and the space-time interval I^2 ? In particular, may we now conclude that these two relations are interdefinable? As a first step towards considering this question, let us consider the relation between symmetry groups and definability in a more general setting.

Let $Th(R, S)$ be a theory containing just two relation-terms R and S . Then a possible model of $Th(R, S)$ will be a structure $\langle A, R, S \rangle$, where A is a nonempty set and R and S are relations on A of the appropriate types. Now let $M = \langle A_1, R_1, S_1 \rangle$ be a model of $Th(R, S)$, and consider the reduced structures $K = \langle A_1, R_1 \rangle$ and $J = \langle A_1, S_1 \rangle$. Both K and J have their associated symmetry groups $Sym(K)$ and $Sym(J)$. Suppose now that these two groups are *not* identical. Then R and S are *not* interdefinable in $Th(R, S)$. This result is an immediate consequence of the following theorem.

Theorem 7.0. Let $Th(R, S_i)$, $i = 1, 2, \dots, n$ be a theory containing the relation terms R, S_i . Let $A = \langle M, R, S_1, \dots, S_n \rangle$ be any model of $Th(R, S_i)$. Then if R is definable in $Th(R, S_i)$ we must have $Sym(\langle M, S_1, \dots, S_n \rangle) \subseteq Sym(\langle M, R \rangle)$. (Here $Sym(\langle M, R \rangle)$ and $Sym(\langle M, S_i \rangle)$ are the symmetry groups of $\langle M, R \rangle$ and $\langle M, S_i \rangle$, respectively.)

Proof: Suppose $h_0 \notin Sym(\langle M, S_i \rangle)$ but $h_0 \in Sym(\langle M, R \rangle)$. Now consider the structure $\langle M, h_0[R], h_0[S_1], \dots, h_0[S_n] \rangle \equiv A'$. By construction, h_0 is an isomorphism of A and A' ; thus A' is also a model of $Th(R, S_i)$. Since $h_0 \in Sym(\langle M, S_i \rangle)$, $h_0[S_i] = S_i$; but $h_0 \notin Sym(\langle M, R \rangle)$, so $h_0[R] \neq R$. Thus $A' = \langle M, R', S_1, \dots, S_n \rangle$, where $R' = R$. Hence, by Padoa's method, R is not definable in $Th(R, S_i)$.³⁷

In order to apply this result to the causal definability of congruence, it is only necessary to note that the mappings in the Extended Poincaré Group are just the pairwise *congruence-preserving* mappings of Minkowski space-time, since *EPG* is just the similarity group of Minkowski space-time. This may be seen as follows. As in affine metric spaces in general, congruence of point-pairs in Minkowski space-time may be defined in this way:

Definition 7.1. Let $A = \langle u, v \rangle$ and $B = \langle x, y \rangle$ be pairs of points in M .

Then A is *congruent* to B iff there is an isometry h of M such that $h(u) = x$ and $h(v) = y$.³⁸

A four-placed congruence relation $C(u, v, x, y)$ (read "the pair u, v is congruent to pair x, y " or " u is just as far from v as x is from y ") may now be defined as obtaining just in case $\langle u, v \rangle$ and $\langle x, y \rangle$ are congruent according to the above definition.

Now suppose h is a *similarity* of M (in particular, $h \in EPG$), and $C(u, v, x, y)$. Then we must also have $C(h(u), h(v), h(x), h(y))$, since the similarity h magnifies both $\langle u, v \rangle$ and $\langle x, y \rangle$ by the same factor. Thus it is easy to see that, quite generally, the similarities of a geometry are just those mappings that preserve the congruence relation.³⁹ So, for Minkowski space, the symmetry group of $\langle M, Con \rangle$ is just the Extended Poincaré Group (*EPG*).

We are now in a position to apply the earlier theorem. Let $Th(CB, C)$ be a theory containing the three-place relation term CB (to be interpreted as causal betweenness) and the four-place relation term C (to be interpreted as the point-pair congruence relation), and containing all sentences true of Minkowski space under this interpretation. Is C definable

John A. Winnie

in $\text{Th}(\text{CB}, \text{Con})$? The above theorem now tells us that this will be the case *only if*

(1) $\text{Sym}(\langle M, \text{CB} \rangle) \subseteq \text{Sym}(\langle M, C \rangle)$, when (M, CB, C) is any model of $\text{Th}(\text{CB}, C)$. But, as we have just seen, $\text{Sym}(\langle M, C \rangle)$ is just EPG , and so by Zeeman's theorem

(2) $\text{Sym}(\langle M, \text{CB} \rangle) = \text{Sym}(\langle M, \text{Con} \rangle)$.

Thus Zeeman's theorem establishes a *necessary* condition for the causal definability of congruence in Minkowski spaces.

Naturally, we now ask: is Zeeman's theorem also a sufficient condition for causal definability of congruence? Or, more generally, does the equality of two symmetry groups (as in (2)), both models of a theory Th , insure the interdefinability in Th of the involved relations?

In general, the answer to this question is negative. The smaller the symmetry groups involved, the less likely it is that their identity will be of any logical importance. It is just this fact about symmetry groups that lies behind the failure of Klein's *Erlangen Program* to provide a comprehensive foundation for geometry. Spaces of highly variable curvature have relatively small symmetry groups (in some cases, the trivial group), and so the structure of these groups provides little or no information about the underlying geometry.

However, the Extended Poincaré Group is not the "small" symmetry group of an inhomogeneous space, but a "large" symmetry group of a strongly homogeneous structure: Minkowski space-time. So although the causal definability of congruence does not follow *in general* from a result such as Zeeman's theorem, it does indeed follow *in this instance* owing to the homogeneity of Minkowski space. This may be shown as follows.

First of all, although Zeeman's theorem merely provides us with a causal characterization of the similarity group of Minkowski space-time, let us suppose that we have managed to provide, somehow or other, a purely causal definition of the *isometry* group PG of Minkowski space-time. We may then define congruence as in definition 7.1 above, and we are done. However, it should be noted that even this last step is not entirely trivial, since the homogeneity of Minkowski space enters here in an important way. Consider, for example, the following alternative approach to congruence in metric spaces:

(**) *Definition.* Let $A = \langle u, v \rangle$ and $B = \langle x, y \rangle$ be ordered pairs of

points in M . Then A is *congruent to* B iff there exists a bijection h from A to B which preserves the metric on M .

Notice that this last definition requires only that h map A onto B , rather than *all* of space M onto itself as well. Thus it may serve to define congruence for all metric spaces, even those with trivial symmetry groups. In addition, this last definition of congruence captures the intuitively desirable feature that two figures should not be deemed incongruent solely on the basis of remote inhomogeneities of the space.⁴⁰ Fortunately it can be shown that, for a wide class of geometries—Minkowski space-time included—the two definitions of congruence are equivalent.⁴¹ The proof of this theorem depends heavily on the metrical homogeneity of the spaces involved, and thus the success of the above definition of congruence in terms of the Poincaré Group is not a trivial matter.

It now remains to provide a causal definition of the *isometry group* PG of Minkowski space-time. First of all, Zeeman's theorem tells us that the causality group CG is the similarity group of Minkowski space-time, so the following succeeds as a causal definition of that group (EPG).

Definition 7.1. $\text{EPG} = \text{df. } \{h \mid h: M \rightarrow M, h \text{ bijective, and for all } x, y, z, \epsilon \in M, \text{CB}(x, y, z) \text{ iff } \text{CB}(h(x), h(y), h(z)).\}$

Now we need to recover the isometry group PG from the similarity group EPG . The following method succeeds in solving this problem for all (nonsingular) real affine metric spaces.⁴²

Call the subset of EPG whose members each fix at least two points " $\text{TF}(\text{EPG})$." More precisely,

Definition 7.2. $\text{TF}(\text{EPG}) = \text{df. } \{h \mid h \in \text{EPG}, \text{ and for some } s, y \in M, x \neq y, h(x) = x \text{ and } h(y) = y\}$.

Clearly $\text{TF}(\text{EPG})$ contains only isometries, since it contains only similarities, all of which have ratio 1. Thus $\text{TF}(\text{EPG}) \subseteq \text{PG}$. Now close $\text{TF}(\text{EPG})$ under finite functional composition, labeling the result $\overline{\text{TF}(\text{EPG})}$. That is,

Definition 7.3. $\overline{\text{TF}(\text{EPG})} = \text{df. } \{h \mid h = h_n \circ h_{n-1} \circ \dots \circ h_1, h_i \in \text{TF}(\text{EPG}), i = 1, \dots, n\}$.

Since the composition of two or more isometries is again an isometry, we still have that $\overline{\text{TF}(\text{EPG})} \subseteq \text{PG}$. However, we can now say more: $\overline{\text{TF}(\text{EPG})}$

contains *all* the isometries, i.e., $TF(EPG) = PG$. This is shown in the proof of the following theorem:

Proposition 7.4. Let M^d be any (nonsingular) affine metric space of dimension $d \geq 2$, with Sim the similarity group of M^d . Then $\overline{TF(\text{Sim})}$ is the isometry group Is of M^d .

Proof: The above argument shows that $\overline{TF(\text{Sim})} \subseteq Is$. Suppose then that $h \in Is$. According to the Cartan-Dieudonné theorem (see Snapper and Troyer, 1971, section 69), there exist at most $d + 2$ symmetries⁴³ of M^d such that $h = s_1 \circ s_2 \circ \dots \circ s_{d+2}$. Since any symmetry of an affine metric space pointwise fixes a $d - 1$ dimensional subspace of M^d , and $d \geq 2$, every symmetry is in $\overline{TF(\text{Sim})}$. Since h is a finite functional composition of symmetries, $h \in \overline{TF(\text{Sim})}$. Done.

Hence the Poincaré Group is just $\overline{TF(EPG)}$, and the accuracy of the following definition is established:

Definition 7.5. $PG = \text{df. } \overline{TF(EPG)}$.⁴⁴

Definition 7.1 above may now be used to define the congruence of point-pairs in Minkowski space-time. With this construction, congruence in Minkowski space-time is shown to be ontologically independent of conventions concerning the isochrony of periodic processes and the congruence of transported spatial units. Congruence in Minkowski space-time is reducible to its causal structure.

The availability of the isometry group PG (or even EPG) of Minkowski space-time makes the definability of its linear structure obvious, since these transformations are all linear. Thus, beginning with two point-events x and y , first consider all transformations in EPG that leave x and y fixed. Let $L(x, y)$ be the set of all points that are left fixed by *all* of the above transformations. When x and y are not light-connectible (not $x \lambda y$), $L(x, y)$ will be the straight line in Minkowski space connecting x and y . If x and y are light-connectible ($x \lambda y$), then $L(x, y)$ may be defined (as in section V) as the set of all events z , such that z is light-connectible to both x and y ($x \lambda z, y \lambda z$). Lines are classified as spacelike, timelike, or lightlike, of course, just in case x and y have the corresponding relation. (Of course, any pair of events on a given line may be chosen.)

The above constructions use the similarity group of Minkowski space together with the causality relation $<$, and, for present purposes, this is perfectly appropriate. Let me, however, put forth the following con-

ture, which is of some independent geometrical interest: given *only* the similarity group (EPG) of Minkowski space-time, it is possible to construct the entire space-time geometry *including the order of points within its lines*. For Euclidean and classical non-Euclidean geometries, their similarity group does not suffice; additional axioms of order must be provided.⁴⁵ However, regardless of the truth of this last claim, the preceding constructions clearly suffice for present purposes, and thus the causal definability of the geometry of Minkowski space-time is established.

Epistemological issues are not hereby resolved, for the above constructions are profoundly nonoperational. Space-time pairs are congruent if (and only if) a global map of M onto itself of the right sort exists; but how are *we*, in concrete measuring situations, to determine the existence or nonexistence of such a map? At first sight, the synthetic constructions of congruence outlined earlier might seem to provide a local method, since fortuitous light parallelograms of finite dimension might then be used to determine interval-congruence. But this turns out to be illusory. These constructions depend upon a prior determination of the collinearity (spacelike and timelike) of at least three events. However, the definitions of spacelike collinearity given by Robb and Latzer, and the definition of timelike collinearity given here (section V) are essentially global. Of course, this does not show that a local causal definition is impossible, and it would be of considerable interest to see this matter resolved.

It should not be supposed, however, that even if Minkowski space-time congruence is bound to global causal structure in an essential way, we are thereby left with the conventionality of congruence for all practical purposes, or worse still, the view that even if *STR* congruence is not conventional, we cannot know it to be so. For auxiliary hypotheses may now serve to provide us with a multiplicity of ways to ascertain global causal structure, and with it, space-time congruence.

VIII. Foundations of Space-time Theories

The reducibility of the metric of Minkowski space-time to its causal structure has not gone unchallenged. Thus Grünbaum, endorsing the views of Reichenbach (1957, section 27) and L. L. Whyte (1953), has held that

Using only light signals and temporal succession without either a solid rod or an isochronous material clock, it is not possible to construct ordinary measures of length and time (Grünbaum, 1973, chap. 13, p. 414).

The basis for this claim is a result that emerges in the course of Reichenbach's axiomatization of special relativity. Reichenbach (1969) attempts to define the Minkowski metric by first constructing the inertial frames of special relativity, then showing these frames to be related by the Lorentz transformations, and, finally, defining the space-time interval as the invariant $-(\Delta x^0)^2 + \sum (\Delta x^k)^2$, where the coordinates $\{x^\alpha\}$ are those of any inertial (Lorentz) frame. The construction begins with the causal precedence relation, together with what he calls "real particles," the latter being the trajectories of continuous timelike paths. Reichenbach's postulates now enable him to "define" (conventions are adopted along the way) the class of all frames of reference S'' in which a light signal propagates according to

$$(1) -(\Delta x^0)^2 + \sum (\Delta x^k)^2 = 0 \quad (k = 1, 2, 3).$$

The question now arises: is the class of frames S'' just the class of inertial frames of special relativity?

The answer to this question is negative; for it can be shown that if K is an inertial frame, then any other frame K' with coordinates $\{x^\alpha\}$ related to those in K by a transformation such as the inversion

$$(2) x^{i'} = \frac{x^i}{-(x^0)^2 + \sum (x^k)^2}$$

will also have an equation of form (1) as its description of the propagation of a light signal (cf. Reichenbach, 1957, p. 172). Fortunately, all of these non-Lorentz frames are essentially of the same sort: some of their spatial points are moving with respect to each other so as to preclude certain sorts of light signal communication. Thus, in Figure 8.0, world lines P' and Q are at rest in K' at $x' = 1, y' = z' = 0$, and $x' = 1/2, y' = z' = 0$, respectively. Any signal sent from P' after event e^* fails to reach world line Q . Why then not use such facts, or the fact that the transformations such as (2) above are not one-one, in order to characterize the inertial frames of S'' ?

This problem was touched upon earlier in the proof of proposition 6.15. There we were considering those transformations that preserve light paths "actively," as conformal mappings of space-time onto itself, rather than "passively," as coordinate transformations. However, the main point remains: there is no bijective mapping (or coordinate transformation) of M (Lorentz frame $\{x^\alpha\}$) onto M (another frame $\{x^\alpha\}$) that preserves light-propagation. In this and a number of other ways, inertial paths may be

causally distinguished from accelerated paths, as the constructions of section V above demonstrate. However, as was mentioned there, these constructions are global: they involve reference to all of the space-time M . As a result, the definitions do not provide operational procedures for determining whether or not the defined relations obtain.

This brings us to the source of Reichenbach's and Whyte's reluctance to rule out the accelerated reference frames in terms of their global properties. Whyte admits that global restrictions are admissible *logical* devices, but rejects the resulting distinction by calling it "a fact of no operational significance" (Whyte, 1953, p. 161); Reichenbach, in a similar vein, remarks:

... since no unlimited spaces can be utilized for a decision, this method is not fruitful. We can always describe systems T [of the accelerated sort] that deviate from systems of class I [inertial systems] only outside the space we have at our disposal. (Reichenbach, 1957, p. 173.)

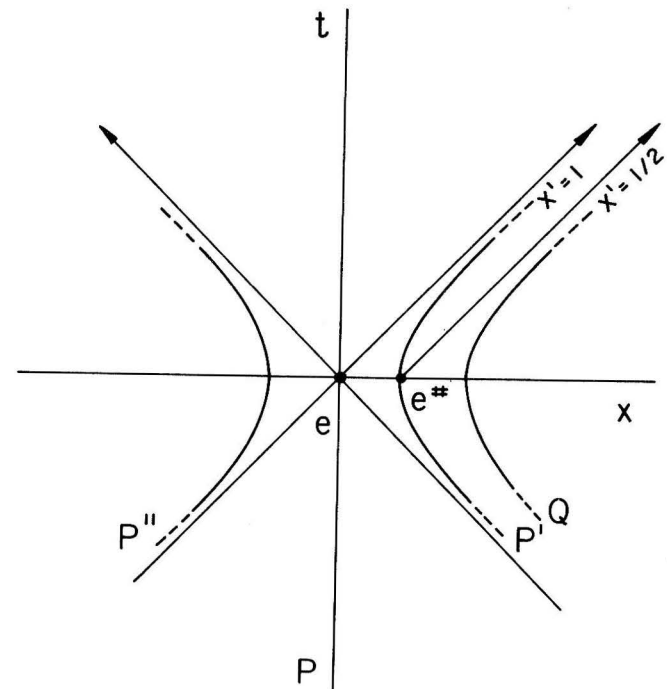


Figure 8.0

Reichenbach's solution is to introduce *rigid material bodies* and use them to determine the accelerations of the points in the unwanted frames. Since the "rigidity" of such bodies is not subject to prior physical determination, it can be only a matter of convention and utility just which class of material bodies is so-taken. Reichenbach realizes that material clocks may be used instead of "rigid" rods, but maintains that essentially the same problem arises all over again for material clocks, now in the guise of attributing congruence to their successive periods. However, as the preceding remarks show, this appeal to 'rigid rods' or 'isochronous clocks' as supplementary to the causal structure of special relativity is needed for purposes of *verification* only, and this is explicitly admitted by Reichenbach (1969, p. 88) as well as by Whyte. Thus writers such as Grünbaum who are concerned to establish the conventionality of Minkowski space-time as a matter of *ontological* fact gain no support for their views from the Reichenbach-Whyte contention. On the contrary, a closer look at the context of Reichenbach's contention reveals that his constructions tend to support, rather than detract from, the causal theory of Minkowski space-time.

Naturally, rejecting operationism, or the verifiability theory of meaning, does not amount to maintaining that global hypotheses about causal structure cannot be confirmed by experiment. To be sure, the results of any set of confirming experiments will also be compatible with a different causal structure in regions not yet investigated, but this is an inductive problem of the sort that inevitably arises when any cosmological hypothesis is under consideration. Furthermore, as the causal theory is extended by adding hypotheses to the effect that light rays travel on causal boundary lines (null lines), free massive particles travel on timelike lines, etc., the ways of obtaining indirect confirmation for global hypotheses of causal structure will multiply indefinitely. In this connection, it is important to note that the equality of the successive periods of atomic clocks has the status of a hypothesized physical fact within the causal theory, as does the congruence of standard rigid rods and clocks under gentle transport. Rods and clocks are devices that survey a region's causal structure, and are accurate or inaccurate to the extent that they yield the causally-grounded congruence classes. Just as space-time has now replaced space and time as the fundamental geometrical object, "etiometry," the measurement of causal structure, replaces and explains both chronometry and geometry.

Still, Reichenbach's failure to distinguish causally the inertial frames of the special theory is only partly explained by his adherence to the verifiability theory of meaning. More fundamental was his failure to recognize the geometrical power and complexity of the causal structure of relativistic space-time. Reichenbach saw clearly that the new causal structure lies behind the relativity of simultaneity, and he saw the demise of absolute *time*; but (unlike Robb) he continued to minimize the differences between classical and relativistic causal structure.⁴⁶

. . . the light-axioms in the theory of relativity do not differ from those in classical theory except for the assertion of the limiting nature of the velocity of light. (1957, p. 175.)

And in referring to the experimental confirmation of his axioms:

Axioms I to II,4 and Axiom II,6 express merely facts known before the theory of relativity. Only axiom II,5 and axiom III are new. These [last] axioms formulate the limiting character of the velocity of light (1969, p. 92).

Unfortunately, one reason so few of Reichenbach's light axioms are new is that not all of them are true. Thus his Axiom II,1 asserts that if we choose an arbitrary event e on any timelike world-line P , and P' is any other timelike world-line, a signal may be sent from e on P to P' . Reichenbach himself recognizes that this is false in relativistic space-times (as P and P' in Figure 8.0 illustrate), but proceeds to use it in any case (1969, p. 31, fn. 9)!

Axiom II,2, however, fares no better. According to this axiom, if e'' is any event on particle world-line P'' and P' is any other world-line, then a signal may be sent from P' , arriving at e'' on P'' . Once again, this axiom is false, as P' and P'' in Figure 8.0 illustrate.⁴⁷

Another aspect of Reichenbach's axiomatization is worth a final note. Being convinced that spatial and temporal congruence were infected with conventions, Reichenbach proceeded to axiomatize relativity theory so as to *exhibit* these conventions clearly. In this way, when statements involving temporal or spatial congruence are derived, the conventions that are presupposed by this assertion may be traced, thereby clarifying the physical content of the original congruence claim. Indeed, by changing *conventions*, and by investigating the effect of these changes on the form of our assertions, we are better able to recognize the "accidental" conventional aspects of a given physical claim.⁴⁸

However, the method has its disadvantages. The most serious difficulty is that the process of conventional variation leads to a set of apparently contradictory statements all having the same physical content. Hence none of them mean what they apparently say. We thereby learn what parts of a locution ought *not* to be taken as an assertion of physical fact, but we may remain as ignorant as ever of that statement's "core" of nonconventional physical content. An appeal to Reichenbach's axioms is of no help here, for, having adopted some particular set of conventions, we find succeeding axioms couched in terms of those conventions. For example, Reichenbach's Axiom V reads:

It is possible to choose the static system relative to A so that the defined spatial geometry will become three-dimensionally Euclidean. (1969, p. 53.)

There is no need to know what a "static system" is in this axiom to appreciate its power *and* its opacity. For the physically interesting question is: *by virtue of what non-convention-laden features of the world is such a choice possible?* The way to exhibit the physical content of our theories is not to *exhibit* their implicit conventions, but to *eliminate* them. Coordinate-free methods are a step in this direction; Robb's axiomatization exhibits the method in its purest form, along with its attendant technical difficulties.⁴⁹

While Reichenbach regards the introduction of metrical conventions as importantly involving epistemological issues, Adolf Grünbaum is most concerned with the ontological status of spatial, temporal, and space-time congruence. For Grünbaum, the philosophical importance of metrical conventions results from their being a symptom of the factual *underdetermination* of physical metrics. The theorist, not nature, forces metric closure by the implicit adoption of physically arbitrary but useful conventions. The congruence of two disjoint spatial, temporal, or space-time intervals is a matter of convention, not because we are unable to ascertain their congruence, but because there *is* no physical relation of metrical equality among the world's spatial, temporal, or space-time intervals.

One moral of this view is clear, and well worth heeding. The standard formulations and practices surrounding a physical theory do not provide clear and explicit guides to its ontological claims. At best, they furnish preliminary clues to be used by the critical scientific realist in his efforts to

create a reconstructed version of the theory that will exhibit its physical commitments with greater perspicuity. In particular, the physical claims of a theory may become especially problematic when the theory is both recent and conceptually revolutionary. For example, it is only by virtue of the relatively recent development of invariant methods that it has become clear (to most relativists) that the relativistic notion of mass is just as much an invariant as was its Newtonian counterpart.

Yet to applaud the merits of Grünbaum's critical approach to the problem of congruence and conventions is not necessarily to agree with his assessments of conventionality or with his diagnosis of its sources. It is well known that Grünbaum locates the source of the conventional ingredients in assertions of spatial, temporal, or space-time congruence in the postulated continuity (locally Euclidean topology) and homogeneity of the underlying manifolds.⁵⁰ However, before such a claim may be established, it is necessary to state explicitly just what relations on a manifold M are to be taken as nonconventional, and why. In a recent work (1973, chap. 16) Grünbaum has attempted to answer the latter question in his general account of the "intrinsicity" of a relation on a set. Only a narrower question will be considered here: namely, given that a set of relations $\{R_i\}$ are held to be "intrinsic" or "objective" on a set M , how are we to assess the physical objectivity of a relation S_0 not in $\{R_i\}$ with respect to the set $\{R_i\}$?⁵¹

Following Klein and Weyl (1949, par. 13), a natural group-theoretic condition, which I propose be regarded as necessary, is the following:

Condition 8.0. Let (M, R_1, \dots, R_n) be such that M is a nonempty set and the R_i are relations on M . Then the relation S on M is objective with respect to (M, R_i) only if the symmetry group of (M, R_i) is a subset of the symmetry group of (M, S) .⁵²

When this condition obtains, S is sometimes said to be an *invariant* of (M, R_i) . Intuitively, the condition amounts to saying that a permutation of M that leaves all of the relations $\{R_i\}$ fixed does not alter the " R_i -objective structure" of M . Should such a permutation then alter another relation S , this would indicate that S brings additional structure to M over and above that already guaranteed by the relations $\{R_i\}$; S is "extrinsic" with respect to (M, R_i) .

We may now apply this condition to the Leibniz space-times of section II. Let (M, CB) be a temporally continuous Leibniz space-time with CB as

its causal betweenness relation. Let d_t be any nontrivial temporal metric on the set of instants of (M, d) and d_s be any nontrivial spatial metric on any of its simultaneity slices. Then, by proposition 2.10, d_t and d_s are not invariant with respect to (M, CB) .

In many cases, Grünbaum contrasts the extrinsicity of the congruence relation on a set M with the relation of 'causal connectivity.' Thus in his discussion of simultaneity in Newtonian mechanics (1969, section 3, subsection 1), he asserts that the arbitrarily fast causal chains of the Newtonian world establish the simultaneity of distant events "as a matter of ordinal temporal fact," whereas "there is no ordinal or topological basis that would yield" the congruences of *successive* temporal intervals. These remarks imply that Grünbaum holds causal connectivity between events to obtain or not as a matter of fact, with the causal relations of a Newtonian world sufficing only for the determination of the temporal order of its events.

As we have seen, the results of section II confirm this *conclusion*; however, they do *not* confirm Grünbaum's diagnosis of the *source* of this causal impotence. For Leibniz space-times, the source is *twofold*: the Leibniz postulate precludes nontrivial *spatial* metrics which are causally invariant, and continuity has nothing to do with this matter. On the other hand, it is the Leibniz Postulate *together with* the postulate of temporal continuity that precludes nontrivial, causally invariant, *temporal* metrics.

From this standpoint it becomes obvious that if we now begin with the relation of causal connectivity γ on *Minkowski space-time*, the question of the objectivity or causal invariance of the congruence relations of this space-time needs to be completely reexamined because of the failure of the Leibniz Postulate in these space-times. The fact that the spatial, temporal, and space-time intervals or paths of special (and general) relativity remain continuous fails to provide even a plausibility argument for their causal noninvariance. Indeed, given Grünbaum's causal nonconventionalism, the causal definability of flat space-time congruence would appear to provide a clear example of what he regards as a nontrivial intrinsic metric.

Furthermore, as we have seen, causal connectivity need not be regarded as merely supplementing a previously supplied topology on M , but we may use the causal connectivity relation to construct that topology as well. In this sense, the causal theory undercuts questions of metrical intrinsicity or extrinsicity by providing a single relation from

which *all* the various "layers" (topology, affine geometry, metric) of the geometry of Minkowski space-time may be constructed. Yet there is a sense in which these affine and metrical structures are not unique, for there are space-times that have neither the same affine nor metrical structure as Minkowski space-time, yet they have the same causal structure. How is this possible?

Suppose (M, g) is a semi-Riemannian manifold with g as its metric tensor field. Then (M, g') is said to be *conformal* to (M, g) just in case $g' = \phi g$, where ϕ is some positive, real-valued, smooth function on M . For our purposes, the important fact about conformal Lorentz space-times is that they have the same null trajectories (see, e.g., Hawking and Ellis (1973), p. 42) and thus the same causal structure. Now if (M, g') is conformal to Minkowski space-time (M, η) , we shall call it *globally conformally flat* (*gcf*). Unless $g' = k\eta$, where k is a constant, (M, g') will not be isometric to (M, η) nor have the same similarity group. But, being conformal to (M, η) , such a nonflat space-time will have the same causal group as (M, η) . The problem posed above is this: how is it possible that the metric η is (up to a constant factor) definable in terms of the causal structure of (M, η) , while there exist semi-Riemannian manifolds (M, g') *not* isometric to (M, η) , but conformal to (M, η) and hence sharing its causal structure?

If we apply the invariance condition C.8.0 above, we obtain the answer to this question at once. First of all, notice that when (M, g') is (globally) conformally flat but not flat, its similarity group will be a proper subgroup of the similarity group (*EPG*) of (M, η) (cf. Levine, 1936), i.e.,

$$(1) \text{Sim}(M, g') \subset \text{EPG} = \text{Sym}(M, \eta) = \text{CG}.$$

Hence there must be a causal automorphism h_0 which is not a similarity of (M, g') ,⁵³ so

$$(2) \text{CG} \not\subseteq \text{Sim}(M, g'),$$

i.e., g' is not invariant with respect to (M, γ) .

It now follows from theorem 7.0 above that g' (up to a constant factor) is not causally definable. As Howard Stein has noted, the problem comes down to this. If $g' = \phi_0\eta$, then a mapping which preserves η up to a constant factor will not, in general, likewise preserve the function $\phi_0: M \rightarrow R^+$. In fact it is easy to see that if ϕ_0 is any nonconstant function on M , *some* function in the extended Poincaré group (*EPG*) will fail to preserve ϕ_0 up to a constant factor. Hence the only way to obtain the invariance of

John A. Winnie

g' is to add a distinguished function ϕ_0 to (M, γ) , obtaining (M, γ, ϕ_0) . Now the metric $g' = \phi_0\eta$ is an invariant of (M, γ, ϕ_0) , and may be defined in a theory $T(\gamma, \phi_0)$ which incorporates an *additional* primitive term ϕ_0 . In terms of the intuitive explanation provided earlier for the invariance condition, the permutation h_0 of M which leaves causality unaltered does change a global conformally flat metric of nonzero curvature, thereby showing the latter to be sensitive to extra-causal features of the space-time structure.

Such cases of the failure of causal invariance in conformally flat spaces lead naturally to consideration of the status of the causal theory with respect to the space-times of general relativity. And if general relativity is considered "liberally," the causal theory is clearly false. For there are models of general relativity, such as Gödel's (1949), such that for every point-event e in M , there is a closed causal curve through e . In such a universe, every event is causally connectible to every other event; hence γ is the universal relation, so the automorphisms of (M, γ) or (M, CB) are just *all* the diffeomorphisms of M onto M . Clearly, no nontrivial semi-Riemannian metric on M is invariant with respect to (M, γ) or (M, CB) , and a causal theory of such space-times must fail.

Suppose, however, that we restrict our attention to relativistic space-times in which the condition of strong causality obtains.⁵⁴ Indeed, the interpretation of general relativity so as to rule out such causality violations was suggested by Einstein in his comments on the Gödel solution (1959). In this case, the failure of the causal theory is not so clear. For example, if we consider all vacuum solutions to Einstein's field equations ($\tau = 0$, where τ is the matter-energy tensor) these reduce to

$$(3) \text{ Ricci} = 0,$$

where 'Ricci' is the Ricci tensor (cf. Schouten, 1954, p. 148). Such spaces are sometimes called "Einstein spaces," or "special Einstein spaces." We now have the following result which generalizes Zeeman's theorem.

Theorem 8.1. If (M, g) is a geodesically complete four-dimensional Lorentz space-time in which $\text{Ricci} = 0$, then either: (1) every local conformal transformation of (M, g) is a similarity or (2) every global conformal transformation of (M, g) is a similarity.

Proof: Case (1): (M, g) is not conformally flat. By a theorem of Brinkman (Schouten, 1954, p. 314), every local conformal transformation is homothetic.

Case (2): (M, g) is conformally flat. Then since $\text{Ricci} = 0$, (M, g) is locally flat (Schouten, 1954, p. 314). But all global conformal transformations of a complete locally flat Lorentz space-time are similarities (cf. Appendix B for an outline of the proof).

Since strong causality ensures that the causality-preserving transformations of a Lorentz space-time be conformal diffeomorphisms,⁵⁵ it follows that the metric of (up to a scale factor) of a large class of relativistic space-times is a causal invariant. The extent to which this result may be generalized poses an interesting question for further investigation. It should be noted that the constructions of Ehlers, Pirani, and Schild (1972), Woodhouse (1973), and Marzke (1964) do not resolve the issue, since they rely on adopting "free particles" in addition to causality.⁵⁶

While the existence of space-times conformal to but not isometric to Minkowski space-time does not impugn the causal theory of Minkowski space-time, the existence of such space-times as models of *general* relativity appears to be a fatal blow to a causal account of these space-times. For if $\mathcal{M} = (M, \eta, \gamma)$ is Minkowski space-time, and $\mathcal{M}^* = (M, \phi^2\eta, \gamma)$ is nontrivially conformal to \mathcal{M} , then $\mathcal{M}, \mathcal{M}^*$ are a pair of models of general relativity having the same causal structure yet metrics which differ nontrivially. They are analogous to a Padoa pair of models in general model theory whose existence suffices to show the nondefinability of the (in this case) metric tensor. Equivalently, the similarity group of \mathcal{M}^* is a *proper* subset of the similarity group of \mathcal{M} and so (by Zeeman's theorem) a *proper* subset of the automorphisms of (M, γ) . Hence there is a causal automorphism of \mathcal{M}^* which is not a similarity of \mathcal{M}^* . Thus the metric $\phi^2\eta$ (up to a constant factor) is not a causal invariant of \mathcal{M}^* . Since invariance is a necessary condition of definability in any reasonable sense of that term, the failure of causal structure to determine the metric of \mathcal{M}^* (up to a constant) would appear to be demonstrated.

It should also be noted that examples like \mathcal{M}^* cannot be ruled out by imposing further restrictive causality conditions on the models of general relativity since \mathcal{M}^* has the same causal structure as Minkowski space-time—itsself a paradigm of good causal behavior. More tellingly, if \mathcal{M} is any admissible model of general relativity, then a conformally equivalent \mathcal{M}^* as above will also be a model, and the argument proceeds as before.

A causal theory of space-time is not necessarily a "relational" theory, as philosophers have used this term. While results such as those dis-

cussed here bear upon the reducibility of space-time structure to causal structure, they do not thereby resolve the issue of the ontological status of these causal relations. Briefly: are events causally related because of the obtaining of relations between material particles or radiation, or do the latter, by their presence, merely modify an independently existing "absolute" causal structure? One reason for regarding standard relativity as endorsing the absolutist interpretation is that there are empty space-times ($T_{\mu\nu} \equiv 0$; "vacuum" solutions to the field equations) which nevertheless have a determinate causal and metrical structure; furthermore, these solutions are not all causally isomorphic or isometric. How can causal or geometrical structure be a function of matter-radiation relations, when there is no matter or radiation present? The difficulty becomes even more striking when it is realized that these empty space-times are not necessarily devoid of "activity." They may, for example, contain pure gravitational waves (plane waves) propagating along some of their null geodesics.⁵⁷

Appendix A. The Two-Point Homogeneity Property of Real Metric Affine Spaces

A discussion of the general structure of real affine metric spaces is given in section VI above. First, the two-point homogeneity property is defined.

Definition A.1. Let $A = (X, (V, \langle, \rangle))$ be a nonsingular, real, affine metric space. Then A is *two-point homogeneous* if and only if for any pairwise distinct points $a, b, c, d \in X$ such that $d^2(a, b) = d^2(c, d)$, there is an isometry $j: X \rightarrow X$ such that $j(a) = c$ and $j(b) = d$.

In other words, for every congruent pair of point-pairs, there is an isometry of the entire space that maps the first pair to the second. The following theorem now asserts that all such spaces (which includes Minkowski spaces) are two-point homogeneous.

Theorem A.1. All nonsingular, real, affine metric spaces are two-point homogeneous.

Proof: Let $A = (X, (V, \langle, \rangle))$, with a, b, c, d distinct, in X , and such that $d^2(a, b) = d^2(c, d)$. First we construct an isometry $i: V \rightarrow V$ such that $i(\langle a, b \rangle) = \langle c, d \rangle$, where $\langle a, b \rangle$ and $\langle c, d \rangle$ are the vectors in V associated (by the action function) with $\langle a, b \rangle$ and $\langle c, d \rangle$, respectively.

Let $\langle\langle \langle a, b \rangle \rangle\rangle$ and $\langle\langle \langle c, d \rangle \rangle\rangle$ be the one-dimensional subspaces of V spanned by vectors $\langle a, b \rangle$ and $\langle c, d \rangle$. Define $i_0: \langle\langle \langle a, b \rangle \rangle\rangle \rightarrow \langle\langle \langle c, d \rangle \rangle\rangle$ by: (1) $i_0(v) \equiv i_0(k\langle a, b \rangle) \equiv k\langle c, d \rangle$, where k is a real constant \neq zero. Clearly i_0 is an isometry, and so by the Witt theorem (Snapper and Troyer, 1971, p. 202) may be extended to an isometry $i: V \rightarrow V$.

Finally, we compose i with a translation $T\langle a, c \rangle$ of A by the vector $\langle a, c \rangle$, obtaining (in the notation of Snapper and Troyer, 1971) $J \equiv T\langle a, c \rangle L(a, i)$. By a standard theorem (Snapper and Troyer, 1971, prop. 381.1), J is an isometry of A . Clearly $J(a) = c$ and $J(b) = d$. Done.

Appendix B. The Causal Group of Locally Flat Space-times

Although a Lorentz space-time (M, g) may be everywhere locally flat, it need not be globally isometric to Minkowski space-time (M, η) . The reason for this is that local flatness does not imply that M has a globally Euclidean topology. Even compact flat space-times are possible, although these are ruled out below by the requirement of strong causality (see section VIII above). It will now be shown that for all complete, flat space-times, every global conformal mapping is a similarity. (A Lorentz space-time is said to be *complete* just in case every affine geodesic may be extended to arbitrarily high (or low) values of its parameter.)

The theory of covering spaces is used freely below.⁵⁸ The general idea is this: the universal covering manifold of a complete, flat space-time $(M, g) \equiv (M, g)$ is just Minkowski space-time (\tilde{M}, η) . Global conformal mappings of M onto itself lift up to global conformal mappings of \tilde{M} onto \tilde{M} in a natural way.⁵⁹ Hence a conformal mapping of M onto M which was not a similarity would lift to an analogous mapping of \tilde{M} (Minkowski space-time) onto itself. But, by Zeeman's theorem, or proposition 6.15 above, there can be no such mapping of \tilde{M} onto \tilde{M} .

To begin with let (\tilde{M}, ρ) be the universal covering manifold of a C^∞ manifold M . Let $\phi: M \rightarrow M$ be a diffeomorphism. We lift ϕ to a diffeomorphism $\tilde{\phi}: \tilde{M} \rightarrow \tilde{M}$ as follows. The function $\phi \circ \rho$ is a smooth mapping from \tilde{M} onto M . Since \tilde{M} is simply connected, by the unique lifting theorem (Massey, theorem 5.1), if $\tilde{x}_0, \tilde{x}_1 \in \tilde{M}$, $x_1 \in M$, and $\phi \circ \rho(\tilde{x}_0) = x_1 = \rho(\tilde{x}_1)$, there is a unique smooth mapping $\tilde{\phi}: \tilde{M} \rightarrow \tilde{M}$ such that $\tilde{\phi}$ is a lift of $\phi \circ \rho$, i.e.,

$$(1) \rho \circ \tilde{\phi} = \phi \circ \rho$$

and $\tilde{\phi}(\tilde{x}_0) = \tilde{x}_1$. That $\tilde{\phi}$ is a diffeomorphism follows from: (a) $(\tilde{M}, \phi \circ p)$ covers \tilde{M} ; and (b), from (1), $\tilde{\phi}$ is a homomorphism of $(\tilde{M}, \phi \circ p)$ into (\tilde{M}, p) . Hence (Massey, lemma 6.7) $(\tilde{M}, \tilde{\phi})$ covers \tilde{M} . Since \tilde{M} is a universal covering manifold, ϕ is a diffeomorphism.

Suppose that $\mathcal{M} \equiv (M, g)$ is a complete, flat Lorentz space-time, and let (\tilde{M}, p) be the universal covering manifold of \mathcal{M} . The manifold \tilde{M} is supplied with a unique metric by using p to lift g to \tilde{M} ; explicitly,

$$(2) \tilde{g}(u_x, v_x) \equiv g(p_*(u_x), p_*(v_x)),$$

where p_* is the differential mapping of p . Since M is flat and complete, so is $\mathcal{M} \equiv (\tilde{M}, \tilde{g})$; but \tilde{M} is also simply connected, and so by the Cartan-Ambrose-Hicks theorem (see Wolf, *Spaces of Constant Curvature* (Boston: Publish or Perish, Inc., 1974), section 1.9) is isometric to Minkowski space-time. Hence we may let $(\tilde{M}, \tilde{g}) = (\tilde{M}, \eta)$, i.e., *the universal covering manifold of any complete, flat, Lorentz space-time is just Minkowski space-time.*⁶⁰

A smooth diffeomorphism $\phi: M \rightarrow M$ is said to be *conformal* just in case

$$(3) g(\phi_*(u_x, v_x)) = e^{\psi(u_x, v_x)} g(u_x, v_x),$$

for some smooth function $\psi: M \rightarrow R$. When ψ is a constant function, ϕ is a *similarity* or *homothetic*. A conformal mapping that is not a similarity is a *properly conformal* mapping; if $\psi \equiv 0$, ϕ is an *isometry*.

We have already shown how to lift any diffeomorphism $\phi: M \rightarrow M$ to a diffeomorphism $\tilde{\phi}$ of \tilde{M} onto \tilde{M} . From (1) and (2) it follows that if $\phi: M \rightarrow M$ is a proper (homothetic, isometric) conformal diffeomorphism of M , then $\tilde{\phi}$ is a proper (homothetic, isometric) conformal diffeomorphism of \tilde{M} . We now have our result at once; for if ϕ were a proper conformal diffeomorphism of M , then $\tilde{\phi}$ would be a proper conformal diffeomorphism of \tilde{M} , Minkowski space-time, and this, by Zeeman's theorem, we know to be impossible. *Hence every conformal mapping of a complete flat space-time is a similarity.* Since for strongly causal space-times, the conformal group and causal group are identical (see section VIII), we have the generalization of Zeeman's theorem: *For every strongly causal, complete, flat space-time, its causal group and its similarity group are identical.*

Notes

1. As in Tarski (1959), where the Euclidean line is asymmetrically ordered by using betweenness and an arbitrary pair of points. In section III below, this procedure is spelled out in detail for Minkowski space-time.

2. Causal connectedness vs. causal connectibility as suitable interpretations of γ will be discussed below. For now, either construal will do.

3. The earlier comments on the *nondirectionality* of causal precedence thus apply derivatively to temporal precedence.

4. Compare this postulate and proposition 2.6 with Leibniz's argument: "And since my prior state, by reason of the connection between all things, involves the prior state of other things as well, it also involves a reason for the later state of these other things and is thus prior to them. *Therefore whatever exists is either simultaneous with other existences or prior or posterior.*" (1956), p. 1083; italics in text.

5. See van Fraassen (1970), chap. 6, section 6.

6. As A. Grünbaum has pointed out (in conversation), we have the following plausible options for the pair (U, γ) :

- (1) U is the set of actual events, γ is 'causal connectedness';
- (2) U is the set of actual events, γ is 'causally connectible'; or
- (3) U is the set of possible events, γ is 'causally connectible.'

The third option is, to my mind, the most plausible, with the physical claim being that the set described in (1) is embeddable in this structure.

7. A formulation of the theory of real order is given in Tarski (1961), p. 214; connectedness (Tarski's Axiom 1') requires the Leibniz postulate.

8. It is important to note, however, that such considerations have clear-cut significance only within the context of a general program (such as the causal theory) of *reduction* for space and time. Often this reduction has been undertaken for operational motives, but this *need* not be its rationale. Reduction may also be desired for the sake of theoretical *explanation*, in which case it is to no avail to reply to the conventionalist that we may *postulate* an intrinsic spatial or temporal metric without scientific embarrassment. Of course; but it is just the postulated metric which is the object of the causal theorist's attempted reduction. It will become clear in the following that such a reduction succeeds provably for special relativity and fails for classical physics.

9. For more on automorphisms, see Weyl (1949), section 13, and also (1952) for developing the right intuitions.

10. Zeeman's theorem and its implications are discussed more fully in section VII below.

11. Throughout this section, I am greatly indebted to the aid of Geoffrey Matthews, Indiana University.

12. In Latzer's interesting paper (1972), chronological betweenness is constructed from the symmetrical relation (λ) of light (null) connectibility.

13. See section IV following for a definition of Alexandroff intervals.

14. A curve in M is here taken to be a smooth mapping from I into M , where I is an open interval of real numbers. A causal curve through e_1 and e_2 is said to go *from* e_1 to e_2 when e_2 has the higher parameter value.

15. See Blumenthal (1970), section 9, for these and other examples.

16. Interesting nonstandard space-time topologies are also possible, as was shown by Zeeman (1967).

17. He was by no means alone in this. See, for example, Lewis (1926), and the admittedly speculative ideas of Bohm (1965). The point is that intervals along a given light path *may* be compared metrically by using an affine parameter along that path, although intervals belonging to distinct nonparallel paths may not be compared meaningfully.

18. Robb (1914) is outlined with proofs omitted in Robb (1921); this is the best introduction to Robb's work. The later Robb volume (1936) is essentially the same as (1914), with some of its constructions and proofs simplified. The introduction to (1936) is a clear statement of Robb's views about the foundations of geometry.

19. A more rigorous formulation is this: let ℓ_t and ℓ_s intersect at e_0 . ℓ_t timelike, ℓ_s spacelike. Then ℓ_t and ℓ_s are orthogonal iff for any event $e_1 \in \ell_t$, $e_1 \neq e_0$, there are events $e_2, e_3 \in \ell_s$ with $e_1 \lambda e_2, e_3$, such that if $e_4 \lambda e_2$ and $e_4 \in \ell_t$ and $e_5 \lambda e_2$ and $e_5 \lambda \ell_t$, then $e_4 = e_5$.

20. Let T_1 be a theory and T_2 an extension of T_1 . Then T_2 may be obtained from T_1 by: (1)

adding postulates which involve no new primitives, or (2) by adding postulates which involve new primitive terms not then definable in T_2 . The above example is of type (2).

21. Not quite a "full" congruence relation in that comparisons are made between (1) timelike intervals, (2) spacelike intervals, and (3) lightlike intervals on the same or parallel light lines. This is more than enough to give us the Minkowski metric.

22. Contrast this with the misleading account given by Reichenbach (1957), section 28.

23. A simple way to see this is to consider an n -dimensional vector space V^n , choose n linearly independent vectors $\langle v_1, \dots, v_n \rangle$, and decree them orthogonal and of unit length. Define their Euclidean inner product as $\langle u^\alpha v_\alpha, w^\beta v_\beta \rangle \equiv \sum u^\gamma v^\gamma w^\gamma = 1, \dots, n$. To obtain another Euclidean inner product differing in its congruence verdict, simply choose, say, $\langle 2v_1, v_2, \dots, v_n \rangle$ as a basis and define the inner product similarly. Clearly v_2 is congruent to v_1 in the first case and not in the second. (An account of metric vector spaces is provided in the following section.)

24. This issue is discussed in more detail in section VI, proposition 6.15, where it is shown that no two Minkowski geometries ($\dim > 2$) with the same causal structure may have metrics that deliver differing congruence verdicts.

25. For the general theory of such spaces, see Snapper and Troyer (1971); this work is used as a reference throughout this section.

26. These conditions are defined as follows: (1) (linearity) $\langle aw + bv, w \rangle = a\langle u, w \rangle + b\langle u, w \rangle$, $\langle w, au + bv \rangle = a\langle w, u \rangle + b\langle w, v \rangle$, (2) (symmetry) $\langle u, v \rangle = \langle v, u \rangle$, and (3) (nonsingularity) if $\langle u_0, v \rangle = 0$, for all v , then $u_0 = \bar{0}$ (the zero vector).

27. In what follows, the fact that the vector space V is real and its inner product nonsingular will be assumed throughout, and thus not always stated explicitly.

28. It should be noted that some writers include positive definiteness in the definition of inner product. The approach taken here follows Snapper and Troyer (1971), chap. 2.

29. The customary formulation of (M) is: there exist basis vectors e_0, e_1, \dots, e_3 of V such that for all $u, v \in V$, $\langle u, v \rangle = -u^0v^0 + u^k v^k$ (sum on $k = 1, 2, 3$), where u^i and v^i are the components of u and v in basis e_0, \dots, e_3 . However, it is easy to show (using the Gram-Schmidt process) that the two formulations are equivalent. The former (M) has the advantage of obvious basis independence. Note that, again, a signature $(-, +, +, +)$ is used.

30. Nothing of importance in what follows depends upon this. Thus we might equally well have defined vectors with equal $|\text{length}^2|$ as congruent.

31. More precisely, let $V(V_m^4)$ be the theory of Minkowski metric vector spaces formulated so as to contain both $\langle _ _ \rangle$ and N (interpreted as 'Minkowski inner product' and 'null vector,' respectively) as primitives. Statement (1) above will now be a theorem of $T(V_m^4)$ showing the definability of N in terms of $\langle _ _ \rangle$. The construction to come shows the definability of $\langle _ _ \rangle$ (under to an arbitrary constant) in terms of N .

32. For generality's sake, a Minkowski space-time is here defined as a four-dimensional differential manifold M , together with a metric tensor field η on M , such that there is a coordinate system $\{x^\alpha\}$ on M onto R^4 in which the components of η are everywhere $\text{diag}(-1, 1, 1, 1)$.

33. More precisely: $TG = \text{df. } \{h \mid h: M \rightarrow M, h \text{ a bijection, and for all } e_1, e_2, e_3 \in M, TB(e_1, e_2, e_3) \text{ iff } TB(h(e_1), h(e_2), h(e_3))\}$; similarly for CG .

34. PG is often called the "inhomogeneous Lorentz group." The Lorentz groups are subgroups of PG , each consisting of rotations about a given point. By virtue of the homogeneity of Minkowski space-time, these groups are isomorphic, hence *the* Lorentz group.

35. For more discussion of the relationship between similarities and isometries in geometry, see Weyl (1949), section 14; and Freudenthal and Bauer (1974), section 22.

36. A magnification of M with center c and ratio r ($r \neq 0$) may be defined as a bijective mapping of M onto itself which (1) has exactly one fixed point c , and (2) maps every other point e' onto the "tip" of vector $k(\vec{c}, \vec{e}')$. For a precise account, see Snapper and Troyer (1971), section 11.

37. For an account of Padoa's method, see Beth (1962).

38. It should be noted that such a definition succeeds only in spaces that are sufficiently

homogeneous to have enough isometries to "move" all congruent pairs onto each other (pairwise free mobility). This property is sometimes called (Birkhoff, 1944) "two-point homogeneity." Euclidean, hyperbolic, and spherical spaces all possess this property. The cylinder is a simple example of a flat space which does not. Nor do Riemannian and semi-Riemannian spaces of variable curvature. However, all real affine metric spaces, and hence Minkowski space, are two-point homogeneous. See Appendix A for a proof of this; also the discussion later in this section.

39. In group-theoretic terms, the similarity group is the normalizer of the isometry group (cf. Weyl, 1949, section 14).

40. Thus consider a Euclidean plane $\langle E, g \rangle$, with g a Euclidean metric tensor field on E . Let A and B be two nearby point-pairs which are congruent (according to both definitions) in $\langle E, g \rangle$. Now choose a new metric tensor g' so as to agree with g on a possibly quite "large" connected region U containing A and B , yet have g' place a single "bump" in $E - U$ somewhere. Isometries of $\langle E, g' \rangle$ must now map the bump onto itself, and thus there may be none which also map A onto B .

41. The theorem, mentioned above, is that all (nonsingular) real affine metric spaces are two-point homogeneous. A proof is given in Appendix A below.

42. For definitions, see the account in Snapper and Troyer (1971).

43. A symmetry of M^d is defined as a reflection of M^d about a fixed hypersurface ($d - 1$ -dimensional subspace) of M^d . See Snapper and Troyer (1971), pp. 219, 386.

44. Professor L. Janos, University of Montana, has pointed out that the involutions of EPC could be used in place of $TF(\text{Sim})$, since every such involution is an isometry and a symmetry.

45. See the work of Bachman for a group-theoretic approach to the classical geometries, especially Bachman, Pejas, Wolff, and Bauer (1974).

46. Lest this judgment seem overly harsh, it should be noted that Robb's initial confidence in the causal theory seems to have been based on the unwarranted conviction that scientific realism *alone* demanded the truth of the causal theory! (Robb, 1921, p. v. ff.)

47. In his footnote commenting on Axiom II, 1 he writes: "There are singular cases in which this axiom does not hold; cf. the example on p. 80. Axiom II, 2, however, holds without exception. This fact indicates a fundamental difference between the two axioms." (1969, p. 31, note 9).

48. Perhaps the best example of the method is its application to the conventionality of simultaneity within the inertial frames of special relativity. For recent analyses, see Grünbaum (1969), Salmon (1969), Winnie (1970), and the critical article by Friedman in this volume.

49. An analysis of length-contraction and time-dilation in special relativity which *eliminates simultaneity* conventions is given in Winnie (forthcoming).

50. Cf. Grünbaum (1968), chap. 3, par. 2.9, 2.10; Grünbaum (1973), chap. 16, is a more recent and detailed account.

51. "Relation" is here used in the broad sense to include functors, sets of relations proper on M , etc.

52. Symmetry groups are explained more fully in section VII above.

53. In the general context of semi-Riemannian spaces, h is *homothetic* or a similarity of (M, g) iff $h^*(g) = kg$ ($k = \text{constant}$), where h^* is the differential of map h (cf. Hicks, 1971, pp. 9, 73). Similarities preserve arc length up to a constant factor.

54. See section IV, p. 45 above for a rough characterization of this condition and a reference.

55. Cf. Hawking, King, and McCarthy, *Orange Aid Preprint*, OAP-405 (1975), or David Malament's dissertation referred to in the next note.

56. For an excellent critical survey of these constructions, see Grünbaum (1973, chap. 22). For some additional results along these lines, see chap. 3 of David Malament's (Rockefeller University) doctoral dissertation (1975).

57. Such cases and others are discussed penetratingly and in detail by Adolf Grünbaum in his contribution to this volume.

58. For the basic theory and relevant results, see, e.g., R. Geroch, "Topology in General Relativity," *Journal of Mathematical Physics* 8, no. 4 (1967):782-786.
59. I am indebted to Professor John Ewing, Department of Mathematics, Indiana University, for helpful comments.
60. It is assumed that, by definition, all Lorentz space-times are connected.

References

- Bachman, F., W. Pejas, H. Wolff, and A. Bauer. (1974). "Absolute Geometry," in Behnke, H., F. Bachman, K. Fladt, and H. Kunle, eds., *Foundations of Mathematics: II: Geometry*. Cambridge, Mass.: MIT Press.
- Behnke, H., F. Bachman, K. Fladt, and H. Kunle, eds. (1974). *Foundations of Mathematics: II: Geometry*. Cambridge, Mass.: MIT Press.
- Beth, E. W. (1962). *Formal Methods*. Dordrecht: Reidel.
- Birkhoff, G. (1944). "Metric Foundations of Geometry, I," *Transactions of the American Mathematical Society*, vol. 55, pp. 465-495.
- Bishop, R. C. and S. I. Goldberg. (1968). *Tensor Analysis On Manifolds*. New York: Macmillan.
- Blumenthal, L. (1970). *Theory and Applications of Distance Geometry*. 2nd ed. Bronx, N.Y.: Chelsea.
- Bohm, D. (1965). "A Proposed Topological Formulation of the Quantum Theory," in I. J. Good, ed., *The Scientist Speculates*. New York: Capricorn Books.
- Earman, J. (1972). "Notes on the Causal Theory of Time," *Synthese*, vol. 24, pp. 74-86.
- Eddington, A. S. (1959). *Space, Time, and Gravitation*. New York: Harper.
- Eddington, A. S. (1965). *The Mathematical Theory of Relativity*. 2nd ed. Cambridge: Cambridge University Press.
- Ehlers, J., F. A. E. Pirani, and A. Schild. (1972). "The Geometry of Free Fall and Light Propagation," in L. O'Raifeartaigh, ed., *General Relativity*. Oxford: Oxford University Press.
- Einstein, A. (1959). "Reply to Criticisms," in P. Schilpp, ed., *Albert Einstein: Philosopher-Scientist*. New York: Harper.
- Fine, A. (1971). "Reflections on a Relational Theory of Space," *Synthese*, vol. 22, pp. 448-481.
- Freudenthal, H. and A. Bauer. (1974). "Geometry—A Phenomenological Discussion," in Behnke, H., F. Bachman, K. Fladt, and H. Kunle, eds., *Foundations of Mathematics: II: Geometry*. Cambridge, Mass.: MIT Press, pp. 3-28.
- Geroch, R. (1970). "Domain of Dependence," *Journal of Mathematical Physics*, vol. 11, pp. 437-449.
- Gödel, K. (1949). "A Example of a New Type of Cosmological Solution of Einstein's Field Equations of Gravitation," *Review of Modern Physics*, vol. 21, pp. 447-450.
- Grünbaum, A. (1968). *Geometry and Chronometry in Philosophical Perspective*. Minneapolis: University of Minnesota Press.
- Grünbaum, A. (1969). "Simultaneity by Slow Clock Transport," *Philosophy of Science*, vol. 36, no. 1, pp. 5-43.
- Grünbaum, A. (1973). *Philosophical Problems of Space and Time*. 2nd ed. Dordrecht: Reidel.
- Haantjes, J. (1937). "Conformal Representations of an n-dimensional Euclidean space with a Non-definite Fundamental Form on Itself," *Koninklijke Nederlandse Akademie van Wetenschappen. Proceedings. Series A, Mathematical Sciences*, vol. 40, pp. 700-703.
- Hawking, S. W. and G. F. R. Ellis. (1973). *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- Hicks, N. (1971). *Notes on Differential Geometry*. London: Van Nostrand.
- Hocking, J. G. and G. S. Young. (1961). *Topology*. Reading, Mass.: Addison-Wesley.
- Klein, F. (1939). *Elementary Mathematics From An Advanced Standpoint: Geometry*. New York: Dover.
- Kronheimer, E. H. and R. Penrose. (1967). "On the Structure of Causal Spaces," *Proceedings of the Cambridge Philosophical Society*, vol. 63, pp. 481-501.
- Latzner, R. (1972). "Nondirected Light Signals and the Structure of Time," *Synthese*, vol. 24, pp. 236-280.
- Leibniz, E. W. (1956). "The Metaphysical Foundations of Mathematics," in L. Loemker, ed., *Philosophical Papers and Letters*, vol. II, Chicago: University of Chicago Press, pp. 1082-1094.
- Levine, J. (1936). "Groups of Motions in Conformally Flat Spaces," *Bulletin of the American Mathematical Society*, vol. 42, pp. 418-422.
- Lewis, G. N. (1926). "Light Waves and Light Corpuscles," *Nature*, vol. 117, pp. 236-238.
- Marzke, R. and J. Wheeler. (1964). "Gravitation as Geometry-I," in Chiu and Hoffman, eds., *Gravitation and Relativity*. New York: Benjamin.
- Noll, W. (1964). "Euclidean Geometry and Minkowskian Chronometry," *American Mathematical Monthly*, vol. 71, pp. 129-143.
- Penrose, R. (1972). *Techniques of Differential Topology in Relativity*. Philadelphia: Society for Industrial and Applied Mathematics.
- Reichenbach, H. (1969). *The Axiomatization of the Theory of Relativity*. Los Angeles: University of California Press.
- Reichenbach, H. (1957). *Philosophy of Space and Time*. New York: Dover.
- Riemann, B. (1973). "On the Hypotheses which Lie at the Basis of Geometry," in, e.g., C. W. Kilmister, ed., *General Theory of Relativity*. Oxford: Pergamon, pp. 101-122.
- Robb, A. A. (1914). *A Theory of Time and Space*. Cambridge: Cambridge University Press.
- Robb, A. A. (1921). *The Absolute Relations of Time and Space*. Cambridge: Cambridge University Press.
- Robb, A. A. (1936). *The Geometry of Space and Time*. Cambridge: Cambridge University Press.
- Russell, B. (1954). *The Analysis of Matter*. New York: Dover.
- Salmon, W. (1969). "The Conventionality of Simultaneity," *Philosophy of Science*, vol. 36, pp. 44-63.
- Schouten, J. A. (1954). *Ricci-Calculus*. 2nd ed. Berlin: Springer-Verlag.
- Snapper, E. and R. J. Troyer. (1971). *Metric Affine Geometry*. New York: Academic Press.
- Tarski, A. (1959). "What is Elementary Geometry?" in Henkin, Suppes, and Tarski, eds., *The Axiomatic Method*. Amsterdam: North-Holland.
- Tarski, A. (1961). *Introduction to Logic*. 2nd ed. Oxford: Oxford University Press.
- Trautman, A. (1965). "Foundations and Current Problems of General Relativity," in Trautman, Pirani, and Bondi, eds., *Brandeis 1964 Summer Institute On Theoretical Physics*, vol. 1. Englewood Cliffs, N.J.: Prentice-Hall.
- Trautman, A. (1973). "Theory of Gravitation," in J. Mehra, ed., *The Physicist's Conception of Nature*. Dordrecht: Reidel, pp. 179-198.
- Van Fraassen, B. (1970). *An Introduction to the Philosophy of Time and Space*. New York: Random House.
- Weyl, H. (1949). *Philosophy of Mathematics and Natural Science*. Princeton: Princeton University Press.
- Weyl, H. (1952). *Symmetry*. Princeton: Princeton University Press.
- Whyte, L. L. (1953). "Light Signal Kinematics," *British Journal for the Philosophy of Science*, vol. 4, pp. 160-161.
- Winnie, J. (1970). "Special Relativity Without One-Way Velocity Assumptions, Parts I and II," *Philosophy of Science*, vol. 37, nos. 1 and 2.
- Winnie, J. "Length-Contraction and Time-Dilation in Special Relativity" (forthcoming).
- Woodhouse, N. M. J. (1973). "The Differentiable and Causal Structures of Space-time," *Journal of Mathematical Physics*, vol. 14, pp. 495-501.
- Zeeman, E. C. (1964). "Causality Implies the Lorentz Group," *Journal of Mathematical Physics*, vol. 5, pp. 490-493.
- Zeeman, E. C. (1967). "The Topology of Minkowski Space," *Topology*, vol. 6, pp. 161-170.