Five

Philosophy of Space and Time

John D. Norton

Philosophy of space and time holds a special place in twentieth-century philosophy of science. Einstein advanced his two theories of relativity, the special in 1905 and the general in 1915, at about the same time as the new movement in philosophy associated with the Vienna Circle, logical positivism and logical empiricism was forming. The founders of this movement—especially Carnap, Reichenbach and Schlick—were fascinated by Einstein's work and sought to bring out its philosophical implications. As a result, some of the clearest applications of their general ideas can be found in the philosophy of space and time. Some examples of these applications are the subject of Part 1 of this chapter and are:

1. An application of the verifiability criterion in Einstein's special theory of relativity. (See "verifiability principle," Chapter 3.) Its target is the absolute state of rest of Newton's theory of space and time, (section 5.1).
2. Antirealist claims about the geometry of space and about the simultaneity of distant events. (See "realism," Chapter 3). The geometry of space is claimed to be chosen by convention and not a factual discovery about the real world. Within special relativity, it is claimed that whether two spatially separated events are simultaneous cannot be determined factually but must be stipulated conventionally, as long as the events cannot causally affect one another, (sections 5.2 and 5.3).
3. A reduction of spatiotemporal relations to causal relations in the causal theory of time. (See "reduction," Chapter 8). All talk of "before" and "after" is claimed to be reducible to talk of what can causally affect what, (section 5.3).

Part 2 contains more challenging material. It provides an introduction to the spacetime methods now commonly used in philosophy of space and time. It has been written for nontechnical readers insofar as very little prior mathematical knowledge is presumed. The major elements of spacetime theories are presented via geometrical pictures and physical metaphors so that the major task of the reader is reduced to visualizing these pictures and metaphors. As we will see, these modern methods are themselves of considerable philosophical interest for they embody an automatic method for determining which elements of a spacetime theory can be chosen by convention and which are fixed by reality.

Part 3 contains applications of these methods. They are used to develop two results. The first is due to David Malament. It supports the claim that if one believes the causal theory of time one cannot also consistently believe the thesis of the conventionality of simultaneity in special relativity. The second is the "hole argument," which originated in Einstein's work. It seeks to establish that a strict form of realism about space and time—a "substantivalist" position—leads to indeterminism in very many spacetime theories.

> $E=mc^2$. Boxed text, such as this, contains explanatory mathematical material that can be skipped by a reader less interested in technicalities.

Part I: Basic Questions

5.1 THE PRINCIPLE OF RELATIVITY

5.1.1 Newton's Absolute Space

Prior to Einstein, the dominant view of space and time was embedded within the mechanics of Newton, the most successful of all scientific theories. In the exposition of his mechanics, his *Mathematical Principles of Natural Philosophy* ([1687] 1962), Newton had distinguished absolute from relative spaces. Relative spaces are the spaces of common experience and a relative space is associated with each observer. The relative space of the reader is the space of the room in which the reader sits and its extension outside the room. If the reader is travelling in an airplane, the relative space will be the space of the cabin and its extension outside the airplane. Now the one process can be described in many different relative spaces. Consider a child riding a carousel at a fairground and a proud parent watching outside from the security of a park bench. In the relative space of the parent and bench, the child is orbiting about the central axis of the carousel. However in the relative space of the carousel, the child is at rest and the world spins about him.

The discussion of relative spaces was not merely an entertaining flourish by Newton but an essential preliminary to the development of his mechanics. He was seeking to lay down the laws that would govern the motions of all bodies. The laws were general descriptions of the behavior of bodies and these descriptions were to be given in one or other relative space. His first law, for example, asserted that in the absence of a net-impressed force a body would remain at rest or in uniform motion in a straight line, that is, in inertial motion. However a body that was at rest in one relative space may be moving nonuniformly in another. This is precisely what we visualized in the case of the parent and the child riding the carousel. To which relative space should Newton refer his laws? If they held in the relative space of the observing parent, then they could not hold in the relative space of the carousel and vice versa. Ideally we would choose a motionless space. Which space could this be? It is not the space of the parent, fixed with respect to the earth. If we believe Copernicus, the earth rotates on its axis and orbits the sun. Again the relative space of the sun is not a motionless space. The sun is a star, one of many in our galaxy, and it orbits the galactic center. As long as we try to base the mechanics on the relative space of some definite object such as the earth or sun, then we risk that object moving along with its relative space with respect to some other object. Again, nothing guarantees that this next body might not itself be in motion or come to be in motion with respect to yet another body. An infinite regress threatens.

Newton solved his problem by denying that the relative space of any body was motionless other than by accident. He announced in a Scholium in Book 1 of his *Mathematical Principles* ([1687] 1962) that of all possible spaces there was one that was eternally motionless, independently of any body that might be associated with it, "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable." It was to this absolute space, the ultimate arbiter of motion and rest, that his laws were to be referred.

5.1.2 The Problem of Verification

Newton's solution raised an immediate problem. How was this absolute space to be distinguished from the many other relative spaces? The obvious answer was to turn to the laws of Newton's mechanics themselves. Absolute space could be identified as that space in which Newton's laws held. Consider a region of the universe remote from massive bodies and a test body in it free of impressed forces. If the body's motion was referred to absolute space, according to Newton's first law it would be at rest or moving uniformly in a straight line. Unfortunately this obvious condition fails to pick out a unique space as absolute space. One relative space which satisfies the condition is the one in which the test body is at rest. But there are others. Consider the relative space of an observer moving uniformly in a straight line with respect to the test body. That observer will see the test body in uniform motion in a straight line, so that the observer's relative space will be one in which Newton's first law holds. In fact there are infinitely many such spaces since there are infinitely many such observers possible, each moving at a different velocity with respect to the original test body. Even allowing for the remaining laws, it turns out that the condition that absolute space be that space in which Newton's laws hold fails to specify a unique space as absolute space. The condition picks out an infinite set of spaces, the inertial spaces, which are the relative spaces of a set of observers moving uniformly with respect to one another in inertial motion.

We should not think that this failure was an oversight of Newton that could be remedied by a small addition to his laws. In fact it was crucial to his theory that his laws work equally well in every inertial space. In formulating this theory, he had to reconcile two apparently contradictory assertions. The first was the Copernican hypothesis that the earth was not at rest but it moved at great speed, spinning on its axis and orbiting the sun. The second was the common fact of everyday experience that we earthbound observers could notice no mechanical effect on the earth's surface attributable to this supposed motion. In Newton's theory, these two assertions were reconciled by noting that, even though the earth spins and orbits, the motion of an observer fixed on the earth's surface is very nearly inertial. (The situation is not so different from that of an athlete running around the circumference of a very large circular stadium. Because of the large size of the stadium, a small segment of the track is almost straight, so that for any brief period the athlete is running in almost exactly a straight line. This would not be the case were the athlete to run in a circle of much smaller radius.) Thus at any instant Newton's laws hold to very good approximation in the relative space of an earthbound observer and to this approximation the observer will not notice any motion of the earth. The approximation is a good one. It requires very sensitive measurements to detect deviations of an earthbound observer from inertial motion. The best known example is the Foucault pendulum experiment, which can be found operating in many science museums. Were Newton's laws to hold in just one inertial space, then this reconciliation would collapse, for the motion of the earth is still only approximately inertial. Over time, the earthbound observer slowly migrates from one inertial space to another as the earth completes its daily rotation and annual orbit of the sun. If significantly differing laws of motion were to hold in each inertial space, then these differences would be revealed to an earthbound observer in the course of the migration and indicate prominently the observer's motion.

5.1.3 The Elimination of Absolute Rest

The precarious compromise of absolute space prevailed for over two centuries in spite of the discomfort of many of Newton's critics. What finally brought the issue to a head were the developments in the theories of light, electricity and magnetism of the nineteenth century. That century saw the revival of the theory that a light ray was a wave and that wave turned out to be the oscillation of electric and magnetic fields. In particular, through the work of such physicists as Maxwell, Hertz and Lorentz, light and its fields were pictured as waves in a medium known as the luminiferous ("light bearing") ether. This ether was a medium that was assumed to pervade all space and it provided physics with another preferred state of rest akin to Newton's immobile absolute space.

By the time Einstein advanced his special theory of relativity in 1905, he had come to see that the status of this luminiferous ether was very similar to that of Newton's absolute space. Newton's laws entailed that no mechanical experiment could distinguish inertial motion with respect to absolute space from rest. Correspondingly, a long series of actual experiments in the nineteenth century had failed to detect the earth's motion relative to the ether. The most accurate and best known of these was the celebrated Michelson–Morely experiment of 1887. Moreover Einstein peered into the innermost heart of the Maxwell–Hertz–Lorentz theory and concluded that as far as *observable* magnitudes were concerned, the theory entailed that inertial motion was not distinguishable from rest. His example concerned a magnet and an electrical conductor such as a wire loop. If the magnet moves through the conducting loop, a current will be induced in the conductor. Alternately, if the conductor moves over the magnet, a current will again be induced in the conductor. It does not matter which of the magnet or conductor is at rest. The same *observable* thing happens, the induction of an electric current, whenever there is *relative* motion between the magnet and the conductor and the magnitude of that current is determined solely by the magnitude of the relative motion. Absolute rest plays no role as far as observables are concerned.

This example launched Einstein's famous paper "On the Electrodynamics of Moving Bodies" ([1905] 1952b) in which he unveiled his special theory of relativity. Referring to this example, he continued:

> Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of the mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good [in all inertial spaces]. We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate . . . The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties. . . . (Ibid., 37–38)

With these words, Einstein introduced the principle of relativity to this theory. The principle asserts, in effect, that the laws of mechanics, electrodynamics and optics are to hold equally well in all inertial spaces. (Einstein picks out the inertial spaces indirectly as those "for which the equations of mechanics hold good.") Newton's quest for the one truly immobile space among them is to be abandoned.

Of course Einstein was not consciously applying the verifiability principle (Chapter 3) when he wrote these words. The criterion was not formulated until over twenty years later. However if we are to identify any justifiable, scientific applications of the criterion then this case surely would be included. The absolute states of rest to which he referred—those of mechanics and electrodynamics—had defied all attempts at experimental identification. Moreover the physical theories involved in each actually predicted that the states of rest admitted no observational consequences which would enable their identification. This is the canonical circumstance for which we seek when we apply the verifiability criterion: the postulation of an entity or state of affairs devoid of observational consequences that would admit its verification. We are thereby enjoined to despise the entity or state as an idle metaphysical conception and to banish it from our discourse.

5.2 CONVENTIONALITY OF GEOMETRY

5.2.1 The Rise and Fall of Euclidean Geometry

The geometry developed by the ancient Greeks is one of the great scientific success stories. Here we see a theory of space which reached its mature form in antiquity and survived so well in that form that even within the last 100 years its standard exposition, Euclid's *Elements* (1956), was still usable as a practical text. In the 1910s, an influential school of thought in philosophy of space and time adhered to an ingenious explanation for this success. Followers of the eighteenth-century German philosopher Immanuel Kant held that the geometry of Euclid *must* hold for all our experience. They urged that the spatial organization of our experiences did not come from external reality, the "things in themselves," but were introduced by our minds in the process of organizing our sensations into something intelligible. This circumstance can be clarified with a parable. Imagine a man compelled always to wear rose-colored glasses. Everything he sees will have a rose tint. He can be sure that every object he will ever encounter will have this tint. However this certainty does not reflect anything about external reality. It results from an unavoidable component of his apparatus of perception. So for the Kantians the truth of Euclidean geometry was guaranteed because that geometry was necessarily imposed onto experience by our minds.

An awkward development of the nineteenth century made the necessity of this imposition less than obvious. It was discovered that it was possible to have consistent geometries based on postulates disagreeing with those of Euclid. Thus even if geometry is imposed by our minds onto experience, it ceased to be clear that the geometry imposed necessarily had to be Euclidean and not one of those "non-Euclidean" geometries. The awkwardness became a serious embarrassment when Einstein completed his general theory of relativity in 1915. That theory entailed that the actual geometry of our own space was non-Euclidean, although the deviations observable in our vicinity from Euclidean geometry were very small.

The new philosophies of science emerging at that time retained the basic idea that the Kantians applied widely: some parts of a theory are reflections of reality and some parts are provided by the organizing mind. However, in the new view, that part contributed by the mind was no longer of fixed or necessary character. It was allowed to vary and it could be chosen at whim or as a convention. To continue the above parable, the man still has to wear glasses, but now he can choose freely the color himself. Thus a new theme entered philosophical analyses of scientific theories, the division of the conventional components from those reflecting reality. The debate over the placement of this division survives in part today in debates over realism and antirealism.

Geometry attracted special attention. It was urged by a number of thinkers such as the French physicist-mathematician Henri Poincaré and more insistently by the German philosopher Hans Reichenbach that the choice of a geometry for space is a matter of convention, for there is no independently true geometry in nature. Thus the geometer could choose whether he would work with a Euclidean or a non-Euclidean geometry. Needless to say one would expect such choices to favor Euclidean geometry wherever practical, since Euclidean geometry is both simple and familiar. But this choice, so the conventionalist thesis asserts, is not to be confused with a discovery of the true geometry, for there is no such thing.

5.2.2 An Argument for Conventionality

How is the conventionality of geometry to be established? Einstein ([1921] 1954a, 235–236) develops the simplest and still best known version of the argument for conventionality of Euclidean geometry and attributes the argument to Poincaré. By itself, the argument runs, a geometry G tells us nothing observable about space. Rather it tells us something about certain idealized structures such as rigid rods which do not actually exist. In order to derive observational consequences about real bodies, we need to resort to physical theories P dealing with such topics as elasticity and thermal expansion to correct for the deviations in the real bodies' behavior from the ideal behavior.

For example, imagine that we wish to check Pythagoras's theorem for the case of a large right angle triangle with sides of length 30 and 40 feet enclosing the right angle. According to the theorem, we expect the hypotenuse of the triangle to be 50 feet and, to check this expectation, we construct the triangle using steel tapes stretched between three points. The catch is that we cannot just pin together three tapes of length 30, 40 and 50 feet. Such steel tapes will always sag a small amount no matter how tautly they are pulled, so that we will need tapes slightly longer than 30, 40 and 50 feet to span the true distances. We can always reduce the amount of sagging by pulling harder on the tapes, but this will in turn stretch the tapes by a small amount since steel is elastic. Thus a *very* accurate check of Pythagoras's theorem with steel tapes can only be carried out if one makes a series of complicated corrections to the behavior of the steel tapes, allowing for their deviations from the behavior of ideally rigid measuring rods. These corrections exploit physical theories of the gravitational and elastic deformation of bodies.

In general terms, observational consequences follow only from $G+P$, the conjunction of the geometry G with the physical theories P. This means that we are free to make conventional modifications to the geometry G as long as we modify our other physical theories P accordingly so that the observational consequences remain unchanged. Thus the one set of observational consequences can be accounted for equally by a large number of conventionally chosen geometries.

5.2.3 Universal Forces, Coordinative Definitions and the Heated Metal Plate

It is at first a little hard to visualize how these conventional elements will appear or, for that matter, just what a non-Euclidean geometry might look like. The following illustration will help. It has been used in various forms by Poincaré, Reichenbach and others and has been modified here for brevity. We imagine a large circular metal plate 10 feet in diameter in a Euclidean space. Since a theorem for circles in Euclidean geometry is that

$$\frac{\text{Circumference}}{\text{Diameter}} = \pi$$

the circumference of the plate will be approximately 31.4 feet (see Figure 5.1). These figures translate into the following operations: A rigid one-foot ruler can be laid 10 times across the diameter and roughly 31 times around the circumference. Were the geometry of the space not Euclidean, then in general we would not obtain π as the ratio of the circumference of the circle to its diameter, but some other ratio larger or smaller than π, when we performed these operations.

Now imagine we discover that the disk is heated in the center but remains cool at its periphery. Imagine also that we carry out the measuring operations with a rod that is not rigid but which will expand when heated as real rods do. (For simplicity, we allow the rod to warm to the temperature of the disk each time we lay it down so that it expands accordingly to the temperature at that part of the plate.) It now follows that if the ruler can be laid 31 times around the circumference, then it will not be possible to lay it the full 10 times across the diameter due to the thermal expansion of the rod towards the center of the plate.

These measurements can be used to justify two different geometries for the plate. First we arrive at Euclidean geometry if we correct the measurements for the thermal expansion of the measuring rod by means of an appropriate physical law for thermal expansion. Second we arrive at a different geometry if we make the alternate physical assumption that no thermal correction is needed and that the rod always measures true distance no matter how its temperature varies. This latter geometry will not be Euclidean. We have, in place of the Euclidean result, the new result:

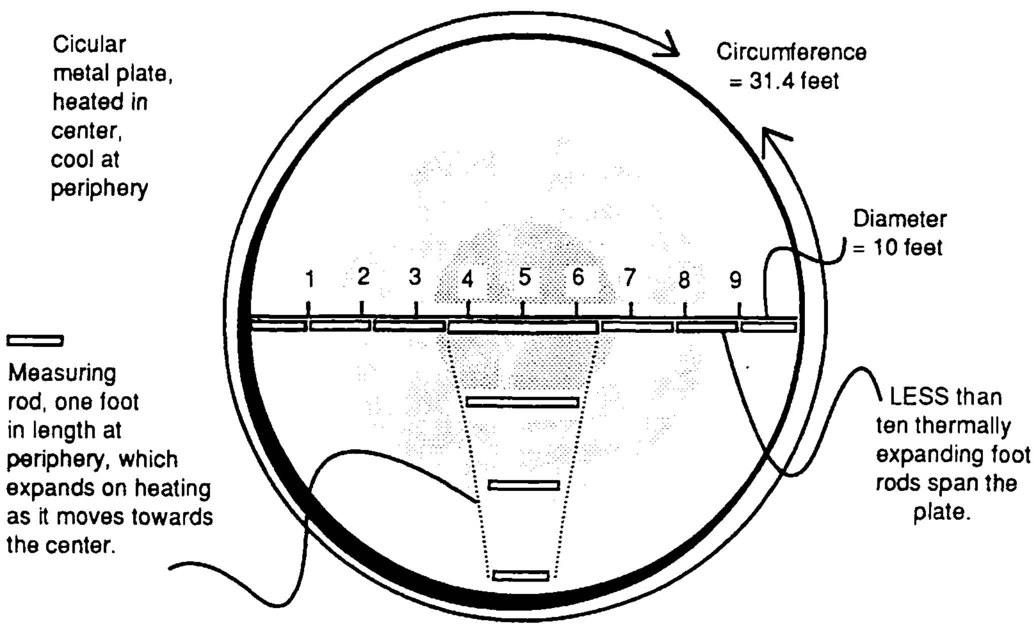

Figure 5.1 Geometry of a heated metal plate.

$$\frac{\text{Circumference}}{\text{Diameter}} > \pi$$

since fewer than ten rod lengths will be needed to span the diameter of the disk. Our choice of geometry then amounts to our choice of which are the rigid rods, the corrected or uncorrected rods. Notice that we *cannot* select the rigid rod by checking for the one that remains constant in length under transport. For what remains constant in length is a matter that can only be known if we already know what is the geometry. We are in a vicious circle. We break it, Reichenbach urges in his influential *Philosophy of Space and Time* ([1928] 1957), by realizing that our choice of rigid rod is a matter of definition. It is a definition that coordinates aspects of physical reality with idealized components of an abstract theory, so Reichenbach calls the definition a "coordinative definition." Now precisely because our choice of rigid rod is a matter of definition, there is no factually correct choice, only conventional preferences. Therefore, since each choice of a rigid rod produces its own geometry, there can be no correct geometry, only our conventional choices.

We would probably be less likely to choose the uncorrected rod as our standard of rigidity (or, better said, congruence of intervals) because we know that various materials expand differently when heated by the same amounts. We might not like to have a material-dependent definition of congruence, especially since the differential thermal expansion of different materials would provide a way of mapping out the temperature changes across the plate, which we might then want to interpret as a cause of length distortion. In the place of thermal deforming forces, Reichenbach considers what he calls "universal forces," which he defines as deforming all materials by the same amount and for which no insulating walls are possible. The situation remains essentially the same as that of the heated metal plate. A universal force will deform our measuring rods, so that the decision to correct or not to correct for such a force amounts to a conventional choice of geometry. In particular the usual assumption that there are no universal forces is a definition leading to a conventional choice of geometry, just as is the decision not to correct for thermal deformation in the case of the heated metal plate.

The notion of universal forces allows Reichenbach to mimic Einstein's figurative "$G+P$" in a concrete way. If we use some physical rod to measure a distance, then its observed length is related to its true length by the equation:

$$\begin{pmatrix}\text{Observed}\\ \text{length}\end{pmatrix} = \begin{pmatrix}\text{True length}\\ \text{according to}\\ \text{geometry } G\end{pmatrix} + \begin{pmatrix}\text{Correction required by}\\ \text{universal force governed by}\\ \text{physical theory } P\end{pmatrix} \quad (1)$$

This equation tells us how to accommodate virtually any conventionally chosen geometry G to a fixed set of observed lengths. We simply choose a universal force whose corrections will make the above equation hold.

The essential point, stressed by the conventionalists, is that some such choice of congruence condition must be made as a definition to break a vicious circle. We cannot know factually what universal forces prevail until we know the geometry; but we cannot know the geometry until we know what universal forces prevail.

Reichenbach actually supplies a more precise version of equation (1), which uses the mathematical apparatus introduced below in Section 5.5. If γ is the metric of a space when universal forces are assumed to be absent, then we can choose any other metric γ' to be the metric of the space as long as there is an object F satisfying

$$\gamma = \gamma' + F \qquad (1')$$

We treat F as the potentials of a universal force field and choose our criterion of congruence to be one which corrects for the deformation due to F. This choice leads us to conclude that the metric of the space is γ'.

5.2.4 Is Everything Conventional?

Can the arguments that were used to establish the conventionality of geometry be used to establish the conventionality of other laws or structures? It seems that they can. To use them to reveal any law as conventionally chosen, all we need is that no observational consequences follow from the chosen law alone. The derivation of these consequences must involve another physical law. We can then make conventional changes to the first law as long as we adjust the second in a way that preserves the observational consequences.

As an example, take Coulomb's law in electrostatics, which describes how one electric charge exerts a force on another. In its field form, it asserts that every electric charge generates an electrostatic field potential φ in the space around it and that this field diminishes in intensity inversely with distance r from the charge, that is, as $1/r$. Now in its field form Coulomb's law *by itself* entails no observable consequences. To recover conclusions in terms of the observable motions of charges, we need the force law, which describes how the field φ acts on other charges. That law asserts that the force exerted on a unit test charge is directly proportional to the quantity "grad φ," which measures the rate at which the field φ changes as we move from point to point in space. But how do we know that the force law is correct? We can only check it independently if we already know what is the field φ. But this we do not know until we have in hand a law, such as Coulomb's, which tells us the fields produced by given charges. We are in a familiar vicious circle. We break the vicious circle with a coordinative definition. We *define* that the field φ is such that the force on a unit test charge is $-$grad φ. Compatibility with observational consequences now leads to Coulomb's law: The field φ diminishes as $1/r$ as we move away from its source. The crucial point is that we are free to choose an alternative coordinative definition.

Examples such as these raise the following questions. Is the conventionality of geometry a position peculiar to geometry and a few other similar fields? Or is it part

For example, we can conventionally replace Coulomb's law with another in which φ weakens as $1/r^3$ with distance r from its source. To preserve the original observational consequences we adjust the force law by defining the force on a unit test charge to be $-\text{grad}\ \sqrt[3]{\varphi}$. Alternatively, we could mimic the introduction of universal forces in geometry in (1) by defining a second field f which mediates the interaction of charges as does the electrostatic field φ. If φ is the electrostatic potential in the case of vanishing f, we can conventionally choose the electrostatic potential to have any other form φ' as long as f is such that

$$\varphi = \varphi' + f$$

is satisfied. The observational consequences remain unchanged provided we adopt a new force law in which the force on a unit test charge is given by $-\text{grad}(\varphi' + f)$. Just as in the case of universal forces in geometry, we have no independent access to the field f so that it is a matter of convention as to whether we set it to zero or not.

of an indiscriminate antirealism in which any law is judged conventional if the law fails to entail observational consequences without the assistance of other supplementary laws? The latter view appears to be a version of the Duhem–Quine thesis (see Chapter 3). This thesis states that it is impossible to test the individual laws of a theory against experience. We can only test the entire theory. Any attempt to test individual laws will fail since we can always preserve any nominated law from falsification by modifying the other laws with which it is conjoined when we derive observational consequences from it.

One conventionalist response is to insist that this general antirealism differs from the true cases of the conventionality of geometry and other related cases such as the conventionality of simultaneity to be discussed in the next section. What distinguishes these latter cases is that the conventionality depends on a very small vicious circle that must be broken by a definition since no independent factual test is possible for the individual components of the circle. The Duhem–Quine thesis does not restrict the manner in which we might protect a law from falsification. We might have to do so by a complicated and contrived set of modifications spread throughout the theory. Some of the components modified may be subject to independent test and thus not properly susceptible to conventional stipulation.

Grünbaum (1973, Part 1) provides an alternate escape from the construal of the conventionalist position as a form of indiscriminate antirealism. He bases the conventionalist thesis on a claim peculiar to space. He urges that space has no "intrinsic" metrical properties, the properties that determine the distances between points, so that these metrical properties must be provided conventionally by us as a definition of congruence.

5.3 THE CAUSAL THEORY OF TIME AND THE CONVENTIONALITY OF SIMULTANEITY IN SPECIAL RELATIVITY

5.3.1 Reichenbach's Constructive Axiomatization of Relativity Theory

Spatial geometry is one of a number of aspects of theories of space and time which, it has been urged, can be chosen conventionally. One might well ask what a theory of space and time might look like if it is built in a way that tries to take full account of these various conventions. One of the early and most significant answers comes in the form of Reichenbach's ([1924] 1969) axiomatization of special and general relativity. Axiomatizations usually seek to condense the physical content of a theory into the fewest axioms of the widest possible scope. Since such axioms inevitably employ the highest-level theoretical structures which mix factual and conventional elements, such axiomatizations do not help the conventionalist sort the conventional from factual content of the theory. To this last end, Reichenbach devised the "constructive axiomatization." Its axioms are statements which are as close as possible to immediate sense experience and thus the basic, factual content of the theory. Of course unfiltered sense experience would lead to an entirely unmanageable axiomatization. Thus Reichenbach employed axioms which rely only on lower-level, prerelativistic theories and which are as close as practical to experiential statements. The axioms are couched in terms of such primitives as "events," "real points," "signals," "rods" and "clocks." The theory is built from axioms such as (from Reichenbach 1924, 29)

> Axiom 1,1. There is no signal chain such that its departure and its return coincide at [a real point] P.

In the course of developing the full theoretical structure of relativity theory, Reichenbach found that he had to supplement his axioms, which contained the theory's experiential content, with definitions of the form of the "coordinate definitions" discussed above in 5.2.3. These definitions comprise the conventional content of the theory and present an integrated picture of the interplay of factual and conventional elements in relativity theory.

5.3.2 The Causal Theory of Time: The Reduction of Time to Causation

An idea implicit in Reichenbach's axiomatization of relativity was what Reichenbach (1956) later called the "causal theory of time." The theory asserts that the temporal order of events is reducible to causal relations between the events. In other words, when we make some assertion about the temporal order of two events, we are really only making assertions about the possibility of one causally affecting the other. Thus Reichenbach ([1928] 1957) offered as a "topological coordinative definition of time order: (p. 136). "If E_2 is the effect of E_1, then E_2 is called later than E_1" (ibid.). More generally, E_2 is later than E_1 if E_1 can causally affect E_2.

Reichenbach assumed that if E_1 and E_2 are causally related we can distinguish the cause from the effect and developed his so-called "mark theory" as one way to support this view. Later workers, such as Grünbaum (1973, Chapter 7), have shown that Reichenbach's attempts here failed and have reconstructed the causal theory on the basis of a symmetric relation of causal connectibility which does not distinguish cause from effect.

The showpiece of the causal theory of time is the possibility of providing an axiomatization of special relativity solely in terms of causal relations so that in some sense the entire space and time of special relativity is reducible to causal relations. The earliest such axiomatization is due to Robb (1914, 1921) in the 1910s and a recent one is due to Winnie (1977).

> A major problem for the theory is the extension to general relativity. As we see in Section 5.8, causal relations in relativity theory depend only on the light cone structure. It has been known since the late 1910s from the work of Kretschmann and Weyl that the specification of the light cone structure in general relativity does not fully determine the complete metrical structure of the spacetime. It would seem that the remaining underdetermined part of the metrical structure cannot be reduced to causal relations. One escape for the conventionalists is to represent the causal structure of spacetime by both light cone and affine structure together.

5.3.3 Conventionality of Simultaneity

One of the conventional definitions contained in Reichenbach's axiomatization of special relativity has become the subject of special attention in philosophy of space and time. It presumes that there is considerable conventional freedom in our determination of which events are simultaneous in special relativity with respect to a given inertial space. (This conventionality should not be confused with the "relativity of simultaneity" in special relativity, which is discussed in Section 5.8 and which involves a change of simultaneity relations with change of inertial space.)

The basic problem is the following. We select two points in an inertial space of special relativity. How are we to judge which events at the first point are simultaneous with which events at the second? This problem is not entirely unfamiliar to us. Imagine that we are celebrating a birthday in two cities *A* and *B* separated by a large distance and that we would like to start lighting the candles on both cakes at the same instant. We could synchronize the two events by means of a long-distance telephone call between the two parties at the time of the lighting. However we might prefer an indirect method. We ensure that identical clocks are located at each party and, prior to the festivities, we synchronize the clocks by means of the long-distance phone call. Once synchronized we can use the clocks to determine which later events are simultaneous at the two places.

If we pare the foregoing procedure down to its barest elements, we have exactly the classic procedure Einstein introduced in his 1905 special relativity paper, "On the

Electrodynamics of Moving Bodies,'' for establishing the simultaneity of distant events. We have identical clocks located at points A and B of an inertial space. We synchronize them by transmission of a signal. We send a signal from the A-clock which is reflected instantly at the B-clock and returns to the A-clock (see Figure 5.2). Were infinitely fast signals possible, we could synchronize the clocks easily. The emission, reflection and return of the signal would all happen instantaneously. So if the emission is arranged for when the A-clock reads 12 noon, we would set the B-clock to 12 noon when the signal arrives. However in special relativity, it is usually assumed that no signal can travel faster than light. Indeed conversations are not transmitted infinitely fast by a telephone system, but at the speed of light. Thus, even if we use the fastest signals available—light signals—there will be a delay at the A-clock between the emission of the signal at A and its return to A. In setting the clocks, we have to decide which event at the A-clock between the emission and the return is simultaneous with the event of reflection of the signal at the B-clock. The situation is shown in Figure 5.3, where the various events at each clock have been spread out in the vertical direction.

If $T_{A\text{-emission}}$ is the time of emission of the signal of the A-clock and $T_{A\text{-return}}$ the time of its return, then Einstein chose *as a definition* that the event exactly half way between the two was the one that was simultaneous with the signal's reflection. This event has the A-clock time coordinate

$$T_{1/2} = T_{A\text{-emission}} + 1/2(T_{A\text{-return}} - T_{A\text{-emission}}) \tag{2}$$

What other choices did Einstein have? The causal theory of time rules out any event at A prior to $T_{A\text{-emission}}$. Such an event could send a signal travelling slower than light to arrive at the reflection event at B, so that the event would be judged as before the reflection. Similarly any event at A after $T_{A\text{-return}}$ would be judged as after the reflection at B. Since the upper limit to the speed of signals is that of light, no event at A between $T_{A\text{-return}}$ and $T_{A\text{-emission}}$ can interact causally with the reflection at B. Therefore any of these events could be chosen as simultaneous with the reflection at B and the clocks synchronized accordingly. The A-clock time of this event is given by

$$T_{\epsilon} = T_{A\text{-emission}} + \epsilon(T_{A\text{-return}} - T_{A\text{-emission}}) \tag{2'}$$

where ϵ must be chosen so that $0 < \epsilon < 1$. The conventionality of simultaneity resides in the conventional freedom to set the value of ϵ anywhere in this interval.

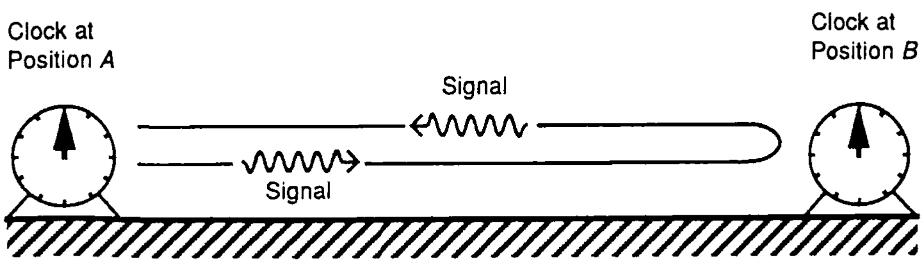

Figure 5.2 Synchronizing clocks by reflection of a signal.

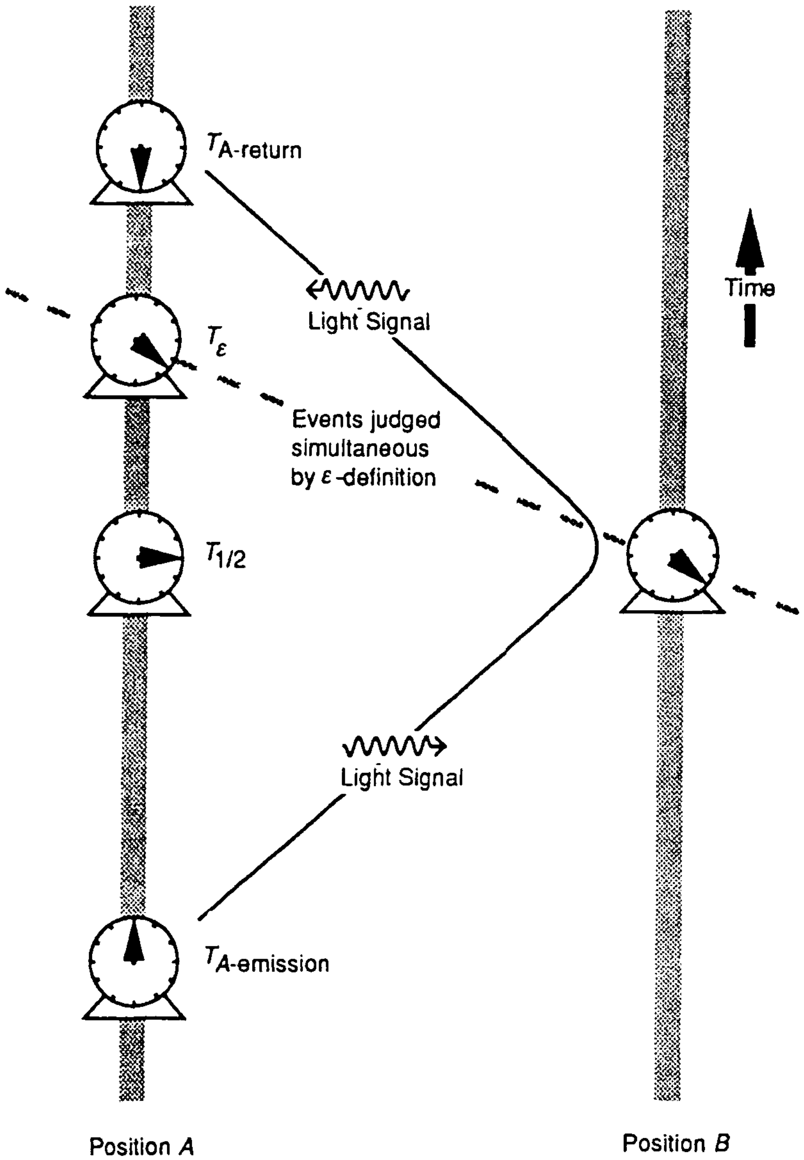

Figure 5.3 Effect of maximum signal speed on synchronization of clocks.

The argument can be strengthened and divorced to some degree from the causal theory of time if it is worked into the familiar form of a vicious circle argument. We could know that the A-event at $T_{1/2}$ is factually simultaneous with the reflection at B as long as we knew that light propagates isotropically, that is, at the same speed in all directions, for then light would take the same time to travel from A to B as from B to A. But to know how fast light travels from one place to another a known distance away, we need to have synchronized clocks at both source and destination so that we can measure the transit time of a light signal and then compute its velocity. Now we can only have *synchronized* clocks if we already know which events at spatially separated places are simultaneous—and the circle is closed. We break this vicious circle by stipulating a value for ϵ. The choice of $\epsilon = 1/2$, "standard" simultaneity, amounts to the stipulation that light travels at the same speed in both directions of the signalling set-up.

If the light signalling method fails to determine a factual simultaneity relation in special relativity, might not some other method? Much of the literature on the subject of conventionality of simultaneity has been devoted to investigating such alternative methods of synchronizing spatially separated clocks and seeking to reveal definitions equivalent to the setting of a value for ϵ in them. See for example Salmon (1977) to get a clear sense that no such convention-free, alternative method is likely to be found. Note that this literature urges the conventionality of the "one-way" velocity light, that is, the velocity between two spatially separated points. The round trip velocity is not taken to be conventional since only one clock at the common source and destination is needed for its measurement.

We return to the conventionality of simultaneity in Section 5.11 to see one of the most dramatic reversals in debates in the philosophy of space and time. David Malament has recently derived a theorem in special relativity which, he urges, shows that the causal relations of special relativity do *not* leave the simultaneity relation underdetermined and thus the relation cannot be set conventionally within the causal theory of time. He shows that the only nontrivial simultaneity relation definable in terms of the causal relations of special relativity is the familiar standard simultaneity relation of $\epsilon = 1/2$.

Part II: Theories and Methods

The purpose of this part is to introduce the methods now used almost exclusively in recent work in philosophy of space and time. These methods differ from those used in Part I in several important ways.

1. There is less emphasis on theories of a space and time as a set of law-like sentences. Rather the theories are approached semantically (see Chapter 3). Thus the activity of the theorist becomes akin to that of the hobbyist model builder, who seeks to represent a real sailboat by constructing a model that captures as many of its properties as possible. The space and time theorist builds models which are intended to reflect the spatial and temporal properties of reality. However the theorist's models are not constructed out of balsa, glue and string, but out of abstract mathematical entities such as numbers.
2. Theories of space and time—including Newton's theory of space and time—are worked into a spacetime formulation. Thus when Newton's theory is compared with its relativistic rivals, all the theories are formulated in the same manner, ensuring that the differences observed are true differences and not accidents of differing formulations.
3. A major theme of Part I was the separation of the conventional or arbitrary elements of a theory from the factual or, as we now say, "physically significant" elements. A means of effecting automatically this separation is built into the notions of "covariance" and "invariance" to be explained here in Part II.