The geometrical foundations of certain relativity theories

By W. H. M'Crea.

(*Received* 26*th November*, 1937. *Read* 3*rd December*, 1937.)
(*Received in revised form* 28*th May*, 1938.)

1. The physical observations that lead to quantitative physical theory are "pointer-readings." The observational data consist of statements to the effect that, when one given set of pointers are incident on certain scale divisions, then another set of pointers are incident on such and such scale divisions. "Pointers" and "scale divisions" are here used in a generalised sense. The question arises as to how it is possible on the basis of a collection of incidence relations of this sort to build up a quantitative theory *i.e.* one involving the concept of measurement. It must be noted that until this is done any numbers associated with scale divisions serve merely as labels.

It is the object of the present note to draw an analogy with the procedure followed in projective pure geometry, which seems to answer this question in principle. This geometry starts from propositions of incidence and, by the well-known process involving the definition of an Absolute Conic or Quadric, can be put into metrical terms. Probably it is generally believed that this type of process provides a foundation also for the introduction of measurement into physics. But, so far as the writer is aware, the only attempt that has been made to follow out such an idea in detail, starting from postulates about what may be regarded as the most elementary physical observations, is that of Robb[1]. His work, however, does not follow a route analogous to the standard one of projective geometry, but proceeds by way of an earlier introduction of parallelism and congruence.

Now should it be the case that an acceptable set of physical postulates were found to be exactly analogous to the axioms of projective geometry, we could take over into physics the whole procedure of the latter theory to show how measurement is introduced.

---

[1] A. A. Robb, *Geometry of Time and Space* (2nd ed., 1936). Possibly the work of A. N. Whitehead (*Principle of Relativity*, 1922) should also be regarded as starting from "projective" foundations and including a treatment of the present problem. However, in his work this problem is not isolated from others concerning the foundations of physics.

Actually no currently accepted set of physical postulates does fulfil this condition. But those of Milne's "Kinematical Relativity[1]" come near to doing so. It is worth while to show how a slight modification of Milne's assumptions will in fact give postulates of the required sort. For the claim of kinematical relativity is that it yields a model universe which reproduces the characteristics of the actual universe sufficiently closely for us to assert that, if we understand how, for example, "gravitation" arises in the model universe, then we understand in principle how it arises in the actual universe. Similarly, if we see how "measurement" arises in a model universe conforming to the modified assumptions, then we may claim to understand in principle how it arises in the actual universe.

We shall find too that the geometrical representation, even though not entirely that appropriate to kinematical relativity, does show why certain types of result should emerge at various stages of that theory.

2. *Postulates of kinematical relativity.* It is necessary briefly to recall Milne's assumptions; we do so first for the case of a single spatial dimension (W. S. §§ 27-35). He postulates the existence of a set of fundamental observers in relative motion in this space. They are supposed to observe each other and explore the space by the emission and reception of light-signals at instants recorded by clocks with which they are equipped (W. S. § 9 *et seq*). It is assumed that light-signals travel in each of two possible senses, say "northwards" and "southwards," in the space, and that each signal possesses some recognisable individuality. Each observer will divide all other observers into two classes, those which he sees "north" of himself and those which he sees "south" of himself, and he is allowed to assume that any signal he emits northwards is seen once, and once only, by each observer north of himself, and similarly for signals and observers in the opposite sense.

These assumptions do not appear in quite these forms in Milne's treatment, but they, and others to be discussed are clearly implied. The last one, for example, corresponds to the fact that in W. S. § 27 Milne assumes that the quantities there called t'_2, t'_3, t_4 all exist and are unique.

It is not essential to the theory to assume that the signals originate at particular observers, and it is more convenient to suppose

[1] E. A. Milne, Relativity, Gravitation, and World-Structure (Oxford, 1935), to be referred to as *W. S.*

that there is an infinite number of signals passing in both senses, and that an observer can make use of these ready-made signals. He might employ a shutter to impress upon them the Morse code suggested by Milne (W. S. p. 27). We then assume that any two opposite signals meet once and once only. The meet of two such signals we shall call an *event*.

In Milne's theory an observer A assigns, in a prescribed manner, to any event in the immediate experience of a second observer B an "epoch" T and a "distance" R. Then a functional relation $R=\phi(T)$ gives A's description of B's motion. Milne, in his second derivation of the Lorentz formulae (W. S. § 32) assumes that for the case of uniform relative motion $\phi(T)$ is a linear function of T, *i.e.* that the uniformity of the motion applies to his conventional distance and epoch. So, leaving aside the question of how A *discovers* that B is an observer in uniform motion relative to himself (see W. S. §§ 23-5), *two* observations suffice to determine for him the whole career of B.

In the general case of three spatial dimensions Milne's fundamental observers are taken to be in uniform motion in this sense. So it appears that he really postulates the existence of a set of observers such that any one of the set can determine the whole career of another by making just two observations on him.

These same postulates may now be stated more formally as follows:

(i) There are infinitely many events.

(ii) There are infinitely many careers of particles in uniform relative motion. (*Events* and *careers of particles* (*or observers*) *in uniform relative motion* are now regarded as indefinable, otherwise than as subjects of these postulates. These particular names are used for the sake of comparison with Milne's theory; they will now be temporarily replaced, for brevity, by *points* and *lines* respectively.)

(iii) Through any point there pass infinitely many lines, and on any line there are infinitely many points.

(iv) Any two distinct points determine a unique line on which they lie, and any two distinct lines determine a unique point which lies on both lines.

(v) A light-signal track is a line (in the sense that it satisfies (iv) and the second part of (iii)); two and only two signal tracks go through each event.

There is however a further assumption in Milne's theory. He supposes that all the careers of particles in uniform relative motion, but not the light-signal tracks, have one and only one event in common, *i.e.* for him the first part of (iii) applies to a single event and not to *any* event. This is the one significant difference between Milne's postulates and those we shall follow here. It is required in Milne's theory in order to make all pairs of his observers *equivalent* in the sense he denotes by $A \equiv B$. This equivalence is not preserved in the present work, but the equivalence which Milne denotes by $A \equiv B$ is preserved, since each of the careers of particles in uniform relative motion is on the same footing as any other, in regard to the whole aggregate. Physically this is important as illustrating the fact that Milne's cosmological principle may be satisfied without his first kind of equivalence of observers being satisfied.

We note now that (i)-(iv) are just the propositions of incidence of two-dimensional projective geometry[1]. We shall add the usual postulates of *order* and *continuity* required to make it a "real geometry" (Baker, *loc. cit.* Chapter 2). The analogous postulates are not given explicitly in Milne's theory, except for events in the immediate experience of an observer (*W. S.* § 13), but they are implied by the fact that he has finally a single real continuum of events.

The resulting two-dimensional space we may call a *space-time plane*. It is now clear, without going into details, that corresponding postulates for the case of three space-dimensions may be written down, and that these will lead similarly to a *space-time four-fold*. The only effective difference from Milne's system will continue to be that we do not take over his assumption of a common event for all the observers in uniform relative motion.

3. *Metrical expression of the geometry.* By the well-known process the properties of the geometrical system may be expressed in metrical terms[2]. Confining attention for the moment to a single space-time plane belonging to the four-fold the metrical expression is introduced with the aid of an Absolute Conic, and any conic in the plane may be selected as the Absolute. In our case we want two particular lines through each point, the signal tracks, to have a special rôle. So we shall take them to be the tangents to the Absolute from the point in question. We recall that we are dealing with a real geometry, so the

---

[1] See for example, H. F. Baker, *Principles of Geometry* I (1929), Chapter I.

[2] H. F. Baker, *Principles of Geometry* II (1930), Chapter 5.

points to which we direct attention must be the points *exterior* to a real Absolute. Then we have as required two and only two tangents through each of these points. It is then known that the geometry is *hyperbolic.* (Baker, *loc. cit.*). Reverting to the four-fold, the corresponding result shows that the metric is expressible in the form

$$ds^2 = k\,\{1 - \tfrac{1}{4}k\,(t^2 - x^2 - y^2 - z^2)\}^{-2}\{dt^2 - dx^2 - dy^2 - dz^2\}. \qquad (1)$$

Here k may be any positive constant. A special form is obtained by replacing ds by $ds/\sqrt{k}$ and letting $k \to 0$; then (1) is replaced by

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \qquad (2)$$

The metric (2) is that of Minkowski space-time, the space-time of special relativity. Letting $k \to 0$ means that we take a degenerate Absolute, in the case of the four-fold a real sphere in the hyperplane at infinity, and in the case of any space-time plane a real point-pair[1], I, J, say.

4. *Physical discussion regarding kinematical relativity.* In the case of the space-time plane just mentioned the signal-tracks through any point are the lines joining that point to I and J. This is just what we require if, in the case of a single space-dimension, we make a unique separation of the two senses (§ 2), such that all observers will agree in labelling any particular light-track as "northward" or "southward." For then all tracks through I, say, are "northward," and all through J are "southward."

In the corresponding case with three space-dimensions we have seen that we have Minkowski geometry. This is the geometry in which we pass from any observer to any other observer by the general Lorentz transformation[2], *i.e.* including where necessary a change of space-time origin. This is therefore the result we should have in Milne's theory were observers other than those having an event in common included, and it is in fact in agreement with the conclusions of Whitrow[3] in regard to such an extension of Milne's work.

On account of the difference from Milne's own case stated above (§ 2) the present work does not yield directly the results of his theory.

---

[1] D. M. Y. Sommerville *Report of the Australasian Assoc. for the Advancement of Science*, XVII. (1924), 140-153. I am indebted to Professor E. T. Whittaker for this reference.

[2] Throughout the work the velocity of light is taken to be unity.

[3] G. J. Whitrow, *Proc. London Math. Soc.* (2), **41** (1936), 529-543.

But it naturally suggests at this stage that he should be led to the geometry in which we pass from one observer to another by the Lorentz transformation without change of space-time origin. This actually is his result (*W. S.* § 36). The metric is given by[1]

$$ds^2 = F\,(t^2 - x^2 - y^2 - z^2)\,\{dt^2 - dx^2 - dy^2 - dz^2\}, \tag{3}$$

where F is an arbitrary function of the argument shown.

A more important conclusion may now be drawn from this discussion. In obtaining (3) Walker was concerned with establishing a formal comparison between Milne's system and the expanding universes of general relativity. He found that the former supplied a counterpart only of those general relativity systems in which the spatial section has constant negative curvature (*i.e.* is a hyperbolic sub-space), and so concluded, "It is clear that Milne has restricted his system by assuming the Lorentz transformation" (*loc. cit.* p. 267).

The geometrical representation shows what is at the basis of this restriction. For the analysis of the assumptions of kinematical relativity in § 2 shows that they do not admit such physical possibilities as that two light-signal tracks should have more than one event in common, that two observers in uniform relative motion should have more than one event in common, or that one of these observers and a signal track should have more than one event in common. These possibilities are excluded in Milne's work, as followed also by Whitrow in his papers on related problems, by the assumptions that the various functions appearing in the analysis are all single-valued. These assumptions are carried over into the geometrical representation by taking the axioms of incidence in the forms we have stated.

The justification for working out the consequences of these assumptions is that they are the simplest from which to start. There is no *a priori* reason to suppose that they apply to the external world, though much may be learned by comparing their consequences with observed properties of the external world. However, general relativity shows that it is possible by a different technique to work consistently without the restrictions involved in these assumptions.

This feature may be illustrated by Einstein's static universe. There the analogues of Milne's fundamental observers are observers fixed in the static coordinate system, and it is easily found that a light-track may go "round the universe" any number of times, and

[1] A. G. Walker, *Monthly Notices, R.A.S.*, **95** (1935), 263-9.

encounter one of these observers every time. Or again two light-tracks may leave a particular observer at one event and, going round the universe by different ways, meet again at a later event in the observer's experience. It does not matter at the moment if, as is probably the case, these particular phenomena do not occur in the actual universe. The point is that such possibilities are not, as in kinematical relativity, necessarily excluded from the theoretical system by the initial assumptions.

There is however further a high probability that phenomena of this general character can occur in local gravitational fields in the actual universe. For it is a result of observation that light rays are bent in passing near the sun. So one should expect that it would be possible to have a body so massive that rays from a point source passing on opposite sides of the body would be bent sufficiently to meet again at some other point, *i.e.* the light-tracks would have more than one event in common. This would correspond to the "lens-like" action of a gravitational field recently discussed by Einstein[1]. At any rate such an effect could not be investigated theoretically by a system like kinematical relativity starts with the implicit assumption it is impossible.

The general relativity expanding universes which Walker finds to be without *counterparts* in kinematical relativity are ones which allow some of the general possibilities just discussed. It is not of course implied that, were the restrictions mentioned removed from kinematical relativity, it would necessarily lead to the same model universes as general relativity. For, apart from this question, the treatments of dynamics and gravitation in the two theories are different.

5. *The space-time derived from the postulates.* We now observe that the metric (1) gives the space-time of the de Sitter universe. Also it can be verified that the lines in the initial projective geometry are the geodesics in its metrical representation, as we expect from the method of introducing the metric. So we have derived the usual "kinematical" properties of this universe from a simple set of pure incidence relations. That this should be possible could indeed have been inferred from the known result that the de Sitter space-time admits an Absolute[2].

---

[1] A. Einstein, *Science*, **84** (1936), 506.

[2] See, for example, Sommerville, *loc. cit.*

We have also derived in a similar way the Minkowski space-time (2). The only "kinematical" difference between the de Sitter and Minkowski space-times is that in the latter, as we have seen, all the observers in a single space-dimension can agree in assigning a "northward" and "southward" sense to it. We may notice also that (1) and (2) are special cases of the form (3) given by kinematical relativity.

A remark must be inserted about the usual exclusion of velocities greater than the velocity of light. It is most conveniently treated after the dynamics of the system has been introduced. But even if we exclude velocities greater than the velocity of light on the ground that they would make it possible for events to be observed before the cause producing them, we are introducing an additional postulate in the denial of this possibility. In Milne's derivation of the Lorentz transformation (*W. S.* § 32) he appears arbitrarily to restrict the relative velocity of the observers to be less than the velocity of light.

6. *Universe with one space-dimension.* In considering a single space-time plane at the beginning of § 3 we explicitly took it to belong to the space-time fourfold, thus implying that the postulates for more than two dimensions (*i.e.* more than one space-dimension) were assumed. This was necessary, for it is well-known (Baker, *op. cit.*, I, p. 120) that the propositions of incidence in a plane only are not sufficient to permit the proof of Desargues's theorem, and thence the development of the subsequent theory leading to the definition of the Absolute.

To follow out the consequences of the assumptions for a single space-dimension only, from the standpoint of this paper, we should require to use the non-Desarguesian geometry of the plane. However, such a geometry does not appear to have been developed in a form suited to this application. So we shall merely note that this geometry is necessarily less restricted than that resulting from the addition of the propositions of incidence for more dimensions. This circumstance should show itself in Milne's work, and this is precisely what we find. For in the case of a single space-dimension Milne finds what he calls the *generalised* Lorentz transformations (*W. S.* p. 41), given in his notation (putting $c = 1$) by

$$T' + X' = p_{12}(T + X), \quad T' - X' = p_{21}(T - X), \tag{4}$$

where $p_{12}(T + X)$ is an arbitrary function of $T + X$, and p_{21} is the function inverse to p_{12}. He later finds (*W. S.* p. 51) that an attempt to extend directly to more dimensions the derivation which leads to

https://doi.org/10.1017/S0013091500002546 Published online by Cambridge University Press

(4) results in inconsistent equations. So the case of more dimensions is more restricted in this respect, as we were led to anticipate. Thus we see in principle why Milne should have obtained generalised Lorentz formulæ for one space-dimension, but did not obtain them in other cases.

In passing it is worth noticing a short way of deriving (4) in Milne's theory. He chooses (*W. S.* p. 29) to define epoch T and distance X *as though* the velocity of light-signals were always constant and equal to c. So the light tracks are given by $T \pm X/c = \text{constant}$. This must apply for all observers, and in his theory these observers agree in the sense ("northward" or "southward") of any signal. Therefore a track $T + X/c = k$, for observer A using T, X coordinates, must be described as a track $T' + X'/c = k'$, for observer B using T', X' coordinates, where k is any constant and k' a constant depending on k. So the transformation from (T, X) to (T', X') must be such that the combination $T + X/c$ transforms into some function of $T' + X'/c$. Similarly $T - X/c$ must transform into some function of $T' - X'/c$. So the transformation must give relations of the form (putting $c = 1$)

$$T' + X' = f(T + X), \quad T' - X' = g(T - X). \tag{5}$$

But for any determination of the functions f, g these two equations must determine the transformation completely.

Now let $X = R$, the distance of B from A as determined by A; then $X' = 0$, and we have from (5)

$$f(T + R) = g(T - R) \, (= T'), \tag{6}$$

giving A's description of B's motion. Again, let $-X' = R'$, the distance of A from B as determined by B (the negative sign occurring since the X, X' coordinates are measured in the same sense while the distances R, R' will be measured in opposite senses); then $X = 0$, and we have from (5)

$$f^{-1}(T' - R') = g^{-1}(T' + R') \, (= T), \tag{7}$$

giving B's description of A's motion.

If then $A \equiv B$ according to Milne's requirement (*W. S.* p. 24) the descriptions given by (6), (7) must be identical. This is the case if

$$g^{-1}(x) \equiv f(x), \text{ so that } f^{-1}(x) \equiv g(x),$$

when (5) becomes

$$T' + X' = f(T + X), \quad T' - X' = f^{-1}(T - X). \tag{8}$$

https://doi.org/10.1017/S0013091500002546 Published online by Cambridge University Press

We have now satisfied all Milne's requirements, so the function f is arbitrary as far as they are concerned, and equations (8) reproduce Milne's equations (4).

The discussion of the one dimensional universe is here mainly of negative character as showing how properties deduced for it are not necessarily capable of extension to more dimensions.

7. *Note on a transformation used by Milne.* The coordinates x, y, z, t in (1), (2) are orthogonal non-homogeneous cartesian coordinates. Corresponding generalised polar coordinates $\mathcal{T}$, χ, θ, ϕ are given by

$$x = \mathcal{T} \sinh\chi \sin\theta \cos\phi,\ y = \mathcal{T} \sinh\chi \sin\theta \sin\phi,\ z = \mathcal{T} \sinh\chi \cos\theta,\ t = \mathcal{T} \cosh\chi. \tag{9}$$

These are the natural coordinates to employ if, as in kinematical relativity, we consider only a set of fundamental observers in uniform relative motion having a unique event in common. For then we take this event as origin, and the world-line of any one of these observers is given by constant values of χ, θ, ϕ. The coordinate $\mathcal{T}$ then measures interval, or *proper-time*, from the origin along the world-line. A given value $\mathcal{T}_0$ of $\mathcal{T}$ specifies equivalent events in the history of all these observers, so that $\mathcal{T}$ may be called the *cosmic time* of the system. These coordinates were introduced from this point of view by Kermack and M'Crea[1].

If we then make the further transformation from $\mathcal{T}$ to τ given by

$$\tau = \mathcal{T}_0 \log(\mathcal{T}/\mathcal{T}_0) + \mathcal{T}_0, \tag{10}$$

the coordinate τ is Milne's *dynamical time*[2] corresponding to the cosmic epoch $\mathcal{T}_0$. It is advisable to keep clear the distinction between the transformation to cosmic time, and the subsequent transformation to dynamical time which is shown by Milne to give his dynamical equations a more familiar form. For then we avoid an apparent difference in this last step between his "non-relativistic"[3] and "relativistic"[4] treatments.

[1] Kermack and M'Crea, *Monthly Notices, R.A.S.*, 93 (1933), 522.

[2] Milne, *Proc. Roy. Soc.*, A., **158** (1937), 177.

[3] Milne, *loc. cit.*

[4] Milne, *Proc. Roy. Soc.*, A., **159** (1937), 171.

Queen's University, Belfast.