

# Einstein's 1935 Derivation of $E=mc^2$

*Francisco Flores\**

Einstein's 1935 derivation of mass–energy equivalence is philosophically important because it contains both a criticism of purported demonstrations that proceed by analogy and strong motivations for the definitions of the 'new' dynamical quantities (viz relativistic momentum, relativistic kinetic energy and relativistic energy). In this paper, I argue that Einstein's criticism and insights are still relevant today by showing how his derivation goes beyond Friedman's demonstration of this result in his *Foundations of Spacetime Theories*. Along the way, I isolate three distinct physical claims associated with Einstein's famous equation that are sometimes not clearly distinguished in philosophical discussions of spacetime theory.

## 1. Introduction

Discussions of the equivalence of mass and energy by historians and philosophers of science seldom include Einstein's 'Elementary Derivation of the Equivalence of Mass and Energy' of 1935 (Einstein, 1935). Yet this paper is of great philosophical interest because it contains criticisms of purported derivations which are still relevant today. Also, rather than introducing the expressions for the 'new' relativistic dynamical quantities, such as relativistic momentum and relativistic kinetic energy, by analogy with their Newtonian counterparts, Einstein shows that stronger motivations can be offered. To establish these claims, I begin, in Section 2, with a brief overview Einstein's first (1905b) derivation by drawing on the work of Stachel and Torretti (1982). My purpose is to reiterate the point that this derivation is *not* fallacious—'purely dynamical' derivations of mass–energy equivalence were not proposed as corrections to Einstein's (1905b). Instead, in Einstein's case, his 1935 derivation is

\* The University of Western Ontario, Department of Philosophy, Talbot College, London, ON, Canada, N6A 3K7

motivated by the desire, often expressed by physicists, to find a derivation of mass–energy equivalence that in no way appeals to Maxwell’s theory of electromagnetism. In Section 3, I outline Einstein’s concerns with derivations that identify the expressions for the relativistic dynamical quantities by analogy with their Newtonian counterparts. Along the way, I show that Einstein is implicitly separating three physical claims associated with  $E = mc^2$  which are not always clearly distinguished.<sup>1</sup> All of this is of current interest because Einstein’s criticisms apply to modern derivations of mass–energy equivalence found in philosophical discussions of spacetime theory. In particular, I show in Section 4 that the derivation of  $E = mc^2$  found in Friedman’s *Foundations of Spacetime Theories* is subject to Einstein’s criticisms. Finally, in Section 5, I show, through a detailed analysis of his 1935 derivation, how Einstein arrives at the definitions of relativistic momentum and relativistic kinetic energy by assuming only the most general form of the mathematical expressions for these quantities and imposing *theoretical* and *methodological* constraints.

## 2. Preliminaries

Einstein first gave a demonstration of the equivalence of mass–energy in his (1905b) paper ‘Does the Inertia of a Body Depend upon its Energy Content’. In this well-known argument, Einstein analyses a thought experiment that consists of a particle emitting two equally energetic pulses of light in opposite directions. Using the transformation equations for the components of the electromagnetic field and for the intensity of the pulse of light which he derived in (1905a), Einstein arrives at his result: ‘the mass of a body is a measure of its energy-content; if the energy changes by  $L$ , the mass changes in the same sense by  $L/9 \times 10^{20}$ , the energy being measured in ergs, and the mass in grammes’ (Einstein, 1905b, p. 71).

For some time, a group of historians and philosophers of science regarded Einstein’s argument as fallacious. For example, in *The Concept of Mass*, Jammer says:

It is a curious incident in the history of scientific thought that Einstein’s own derivation of the formula  $E = mc^2$  [ ... ] was but the result of a *petitio principii*, the conclusion begging the question (Jammer in Stachel and Torretti, 1982, p. 760).

Jammer reaches this conclusion because he believes the only way to justify a step in Einstein’s derivation is to *assume* both the expression for the kinetic energy of a particle *and* the relation between the change in mass of the emitting body and the energy emitted by the body, viz  $m_i - m_j = L$ , where  $m_i$  and  $m_j$  represent the mass of the body before and after the act of emission. If Einstein had indeed made these assumptions, he would be guilty of begging the question. Recently, however, Stachel and Torretti (1982) have shown convincingly that Einstein’s

<sup>1</sup> A notable exception to this approach, which is similar to Einstein’s (1935) argument, is Feynman *et al.* (1963, pp. 16–6–16–10).

original derivation contains no such fallacy.<sup>2</sup> Stachel and Torretti observe that while it is true that Einstein derived the expression for the kinetic energy of an ‘electron’, i.e. a structureless particle with a net charge, in his earlier (1905a) paper introducing special relativity (SR), he nowhere uses this result in his (1905b) derivation. They correctly advise the reader to ‘note how carefully Einstein avoids any assumption about how the relativistic energy depends on the velocity, mass or any other parameters of the body’s internal state’ (1982, p. 761). Stachel and Torretti then show that Einstein’s critics have overlooked two key ‘moves’ available to Einstein. First, one can always switch from the passive to the active interpretation of the Lorentz transformations. Second, one can define the inertial mass of a body using the Newtonian limit. And these ‘moves’ are sufficient to complete a sound demonstration.

Despite the demonstrated cogency of the argument, however, as physicists often remark, Einstein’s original derivation is undesirable because it appeals to Maxwell’s theory. Einstein himself held this view. In the introduction to his (1935) derivation of mass–energy equivalence, he points out that although SR grew out of Maxwell’s electromagnetic equations, SR itself is *independent* of Maxwell’s theory. Einstein wants to show that the results of SR, of which he considered  $E = mc^2$  to be the most significant, are likewise independent of Maxwell’s theory. For Einstein, this becomes particularly important when one faces the possibility that energy is quantised, i.e. when we realise that ‘we do not know the extent to which the energy concepts of the Maxwell theory can be maintained in the face of the data of molecular physics’ (1935, p. 223). Thus, the goal is to find a derivation of the equivalence of mass and energy that depends only on the physical principles that lie at the core of SR.

### 3. Einstein and Derivations that Proceed by Analogy

Einstein’s ‘Elementary Derivation of the Equivalence of Mass and Energy’ begins by asking us to consider the following ‘derivation’ of the equivalence of mass and energy. Suppose a point particle of mass  $m$  travels with velocity  $\mathbf{u} = (u_1, u_2, u_3)$  with respect to some coordinate system  $S$  adapted to an inertial frame  $F$ . We can define the four-velocity of this particle, using  $\gamma(\mathbf{u}) = 1/(1 - u^2)^{1/2}$  to represent the Lorentz factor (with  $c = 1$ ), as the four-vector with components (in this coordinate system):

$$\mathbf{U} = \gamma(\mathbf{u})(1, u_1, u_2, u_3) = \gamma(\mathbf{u})(1, \mathbf{u}). \quad (1)$$

Hereafter I denote  $\gamma(\mathbf{u})$  by  $\gamma$ , when the parameter is obvious from the context. One can then obtain a vector ‘connected’ with the motion of the particle by multiplying  $\mathbf{U}$  times the mass  $m$  of the particle:

$$\mathbf{X} = m\mathbf{U} = (m\gamma, m\gamma\mathbf{u}). \quad (2)$$

<sup>2</sup> As with all demonstrations of mass–energy equivalence, Einstein assumes that all of the mass of a body can be converted into energy.

Using the familiar binomial expansion for  $\gamma$ , and neglecting terms of the third order and higher in the velocity, we obtain the following expression for the components of this vector:

$$\mathbf{X}_A = (m + \frac{1}{2}mu^2, mu_1, mu_2, mu_3). \quad (3)$$

In this approximation, the spatial components of the four-vector  $\mathbf{X}_A$  are equal to the classical expression for the momentum of the particle while the temporal component is equivalent to the classical expression of the kinetic energy plus an additive constant  $m$ . Going back to the exact expression of the vector defined by (2), it seems natural, Einstein observes, to take the spatial components of  $\mathbf{X}_A$  as the relativistic three-momentum and to take the temporal component minus  $m$  as the relativistic kinetic energy and thus define:

$$\mathbf{p}_{\text{rel}} = m\gamma\mathbf{u} \quad (4)$$

and

$$KE_{\text{rel}} = m(\gamma - 1). \quad (5)$$

Einstein now asks: how is one to *interpret* the temporal component  $m\gamma$  of the original vector  $\mathbf{X}$ , ‘the expression of which after all is the really significant one?’ (1935, p. 224). It seems reasonable to interpret this component as an energy term since it is the non-constant term in the definition of  $KE_{\text{rel}}$ . We can then define the relativistic energy of the particle as

$$E_{\text{rel}} = m\gamma. \quad (6)$$

Furthermore, from this definition of  $E_{\text{rel}}$  we see that when the velocity is zero, there is an energy  $m$  associated with the body. We now ascribe this *rest-energy*  $m$  (equal to  $mc^2$  when we factor in the speed of light) to the particle and conclude that the rest-mass of a body and its rest-energy are equivalent. So have we given a proof of the equivalence of mass and energy?

Einstein argues that there are two reasons why the answer to this question is no. First he says:

Of course, this derivation cannot pretend to be a proof since in no way is it shown that this impulse satisfies the impulse principle and this energy the energy-principle if several particles of the same kind interact with one another; it may be a priori conceivable that in these conservation-principles different expressions of the velocity are involved (1935, p. 224).

Einstein is suggesting that we have no grounds to suppose that the quantities  $m\gamma\mathbf{u}$ ,  $m(\gamma - 1)$  and  $m\gamma$  should be *interpreted* respectively as the relativistic momentum, relativistic kinetic energy, and relativistic energy of the particle. Before we can adopt such an interpretation, we must show that these quantities satisfy their respective conservation laws. So, for example, in the theoretical analysis of perfectly elastic collisions involving an arbitrary number of particles, the quantities  $m\gamma\mathbf{u}$  and  $m(\gamma - 1)$  should be conserved. While it seems plausible,

*prima facie* at least, to interpret these quantities as ‘relativistic momentum’ and ‘relativistic kinetic energy’ because they reduce to their Newtonian counterparts, Einstein is pointing out that there may be other quantities that satisfy this requirement. However, not all of *these* quantities will satisfy the conservation principles which we take to be definitive of the corresponding concepts.<sup>3</sup> Thus, having the appropriate Newtonian limit is only a necessary condition for such an interpretation; it is clearly not sufficient. What justification might count as sufficient for Einstein will be made clear in Section 5 where I discuss the rest of his paper.

Einstein’s second reason for saying that the purported demonstration is not a proof reveals the core of what is at issue in demonstrations of  $E = mc^2$ . Einstein first points out that ‘it is not perfectly clear what is meant in speaking of the *rest-energy*, as the energy is defined only within an undetermined additive constant’ (pp. 224–225). What is significant here, which we know from Einstein’s first derivation, is that by *rest-energy* we do *not* mean this additive constant. In both the Newtonian and relativistic case this additive constant is set to zero by convention. Einstein’s point is that it is not clear what we *do* mean. Einstein continues:

in connection with this, however, the following is to be remarked. Every system can be looked upon as a material point as long as we consider no processes other than changes in its translational velocity as a whole. It has clear meaning, however, to consider changes in the rest-energy in case changes are to be considered other than mere changes to the translational velocity (p. 225).

Einstein is pointing out that for any collection of particles (or more generally any mass distribution) we can always consider the entire system as a material point provided we are only considering changes in the translational velocity of the system. We then simply treat the system as a material point, viz the centre of mass of the system, which moves with a given velocity. However, there are cases where it makes sense to speak of a change in the *rest-energy* of a particle. For example, when a particle emits energy, say in the form of electromagnetic radiation, but does not change its state of motion, then surely its *rest-energy* has changed even though there have been no changes to its translational velocity. In this case, interpreting the quantity  $m\gamma$  as the energy of the particle has the consequence that ‘in such a transformation of a material point its inertial mass changes as the rest-energy; this assertion naturally requires a proof’ (p. 225).

At this stage, it is worthwhile to point out that Einstein is implicitly separating three distinct claims associated with his famous equation which are sometimes not clearly separated in philosophical discussions of spacetime theory. The first

<sup>3</sup> Einstein naturally does not make the same demand of the concept of rest-mass—presumably because he does not believe that its conservation, although it is a property of classical mass, is definitive of the concept. Einstein seems to be assuming that the inertial mass of a body is defined as the ratio of force to acceleration.

claim, i.e. the assertion that, according to Einstein, ‘naturally requires a proof’ (p. 225), can be expressed as:

ME1. When there are no changes to a particle’s translational velocity, a change in the rest-energy of the particle results in a proportional change to its inertial mass, and conversely.

It is significant that Einstein relates a change in the rest-energy of a particle to a change in its *inertial* mass; what changes is the coefficient that appears in the expression for the kinetic energy and the momentum of the particle. Einstein *assumes* that these coefficients are the same and claims to later find partial justification for this assumption. An important consequence of ME1 is that when we combine masses by any method that releases energy, or when a particle splits into two or more particles and releases energy, the classical law of conservation of mass will no longer hold.<sup>4</sup>

Once we have secured the definition for the ‘new’ dynamical quantities, one can, from these definitions *alone*, infer the second claim associated with the equivalence of mass–energy:

ME2. If there is a change in the translational velocity  $u$  of the particle relative to a given inertial frame, what we now call the relativistic mass of the particle, and its relativistic energy, change.

One way to describe this, using the Minkowski formulation of SR, is to say that observers will determine different components for the four-momentum of a given particle depending on their relative state of motion. Einstein distinguishes ME2 from ME1 because he considers changes that affect only the translational velocity of the particle separately from ‘other changes’. In this context, ME1 can be interpreted as saying that for a given observer, i.e. with respect to the *same* inertial frame, during an inelastic collision between two particles (i.e. when we allow changes not just of translational velocity but also to the original masses and rest-energies of the particles) the change in the rest-mass of a particle is equal to the change in rest-energy. Conversely, if a particle loses energy through some physical process, the change in energy is equivalent to the change in its rest-mass. This is *not* a claim about how different observers ‘decompose’ a four-vector into different spatial and temporal components. The claim is that with respect to a single frame of reference, if the particle loses energy, then the coefficient of the kinetic energy (i.e. the particle’s inertial mass) changes. Finally, as Mermin and Feigenbaum (1990) have pointed out, this is different from a third claim associated with the equivalence of mass–energy,

<sup>4</sup> Since one can define collisions quite generally to include these kinds of changes, one can say that during inelastic collisions the inertial mass of a particle is no longer conserved in relativistic dynamics.

usually stated as:

ME3. The rest-energy of a particle is equal to its rest-mass  $m_0$  (times the speed of light squared).

ME3 states that when the velocity of a particle is zero with respect to some inertial frame, the energy which must be attributed to this particle relative to that frame is  $m_0c^2$ . This attribution is entirely conventional because, as Einstein points out, one can always choose the zero-point of the energy arbitrarily — though some choices recommend themselves largely for computational simplicity.

In sum, Einstein's concerns are as follows. First, we should not interpret  $m\gamma$  as the relativistic *energy* of a particle of mass  $m$  moving with velocity  $u$  until we have shown that, theoretically at least, this is a quantity that is conserved in all collisions. (Of course, it does not follow from this that this quantity will *actually* be conserved when we perform experiments.) To adopt such an interpretation is 'unjustified' because we do not yet know that this quantity meets one of the defining characteristics of the concept of energy. Second, whatever the expression for the *rest-energy* turns out to be, there are clear cases where it makes sense to speak of the rest-energy of a particle changing, e.g. when a particle releases energy (say in the form of electromagnetic radiation), but where there is no change in its state of motion relative to some inertial frame. *If* we interpret  $m\gamma$  as the energy, this has the consequence that when the rest-energy changes, then the particle's rest-mass changes in the same way (and conversely). This is a new physical result that requires a 'proof'. One should be able, in other words, to demonstrate that this result follows from the accepted physical principles and definitions that lie at the core of the theory. It should not be a consequence of an 'unjustified' interpretation of a physical quantity.

#### 4. Friedman and Mass–Energy Equivalence

Einstein has pointed out a significant lacuna in derivations that proceed merely by analogy. While these derivations may be of great heuristic value, they often fail to separate the three distinct physical claims associated with the equivalence of mass–energy and, in the context of philosophical discussions of spacetime theory, sometimes fail to establish ME1. For example, Friedman's derivation of mass–energy equivalence, in his *Foundations of Spacetime Theories* (1983, p. 142 ff.), falls short of providing the kind of strong motivation for the definitions of the 'new' dynamical quantities that Einstein's (1935) derivation provides. Furthermore, presumably because relativistic dynamics is not his principal concern, Friedman stops short of proving the claim ME1, which is the core of Einstein's result. We can better appreciate the extent to which Einstein's concerns are still relevant by taking a closer look at Friedman's derivation.

Friedman begins by demonstrating the consequences to the laws of dynamics of changing from Newtonian absolute time to the proper time of SR.<sup>5</sup> First, he asks us to consider the law of motion for Newtonian dynamics, i.e.  $F = ma$ , which he writes as:

$$m\zeta^a\nabla_a\zeta^b = F^b, \quad (7)$$

where  $\zeta^a$  is the tangent vector field to some curve  $\sigma(t)$  parametrised by the absolute time  $t$  and  $F^b$  is a spacelike vector field.<sup>6</sup> In Newtonian spacetime, this implies that  $F^b$  lies on the unique plane of simultaneity through a point  $p$  for each  $p$  on  $\sigma(t)$ ;  $t$  is any time function, i.e. for any displacement four-vector  $\zeta^a$ ,  $t_a(\zeta^a) = t$  where  $t$  is the magnitude of the projection of  $\zeta^a$  onto a time axis;  $m$  is the Newtonian inertial mass of a particle. Friedman explains the requirement that  $t$  (in equation (7)) must be a time function by providing the following result. Let the curve  $\sigma(t)$  be re-parametrised by an arbitrary parameter  $s$ . The requirement that  $F^b$  must be a spacelike vector, together with the requirement that the temporal ‘metric’  $t_a$  is compatible with the derivative operator, i.e.  $\nabla_a t_b = 0$ , imply that  $s$  is a ‘suitable parameter’ in the Newtonian law of motion if, and only if,  $s$  is itself a time function, i.e. if  $s$  is a linear function of  $t$ .

Friedman then writes down the ‘corresponding law of motion for relativistic dynamics’ (Friedman, 1983, p. 143):

$$m_0\zeta^a\nabla_a\zeta^b = F^b, \quad (8)$$

where now  $\zeta^a$  is the tangent vector field to a curve  $\sigma(\tau)$  parametrised by the proper time  $\tau$ , and  $m_0$  is the rest-mass of the particle, i.e. ‘the mass that would be determined by an ideal measuring device co-moving with the particle’ (Friedman, 1983, p. 143). Again,  $F^b$  is a spacelike vector orthogonal to  $\zeta^a$ . Friedman then notes that the ‘chief difference’ between equations (7) and (8) (which is concealed in our notation) is that we have changed the scalar parametrising the curve  $\sigma$  from the absolute time  $t$  to the proper time  $\tau$ : ‘proper time has replaced Newtonian absolute time’ (Friedman, 1983, p. 143).

In the Newtonian case, Friedman observes, the requirement that  $F^b$  be a spacelike vector implies that it lies on the unique plane of simultaneity through some point  $p$  along the curve  $\sigma(t)$ . In the relativistic case, one cannot demand this since there is, of course, no unique plane of simultaneity through  $p$ . There is,

<sup>5</sup> Throughout this discussion I assume that the reader is familiar with the modern way of presenting spacetime theories using differential geometry on manifolds. I have also taken the liberty to translate Friedman’s notation into standard abstract index notation. Also, I am assuming the standard distinction between a *curve*, which is mapping  $\sigma: I \subseteq \mathbb{R} \rightarrow M$  (where  $\mathbb{R}$  is the set of real numbers,  $I$  is an interval of  $\mathbb{R}$ , and  $M$  is a manifold) and a *path*, which is the set of points which is the image of this mapping on the manifold, i.e.  $\sigma[I]$ .

<sup>6</sup> The vector field  $F^b$  is spacelike because in the context of Newtonian spacetime the four-acceleration is always ‘orthogonal’ to the four-velocity. This is also the case of Minkowski spacetime. The four-velocity is always timelike, so the four-acceleration must be spacelike.

however, a unique plane of simultaneity through  $p$  that is orthogonal to  $\sigma$ . This is the plane of simultaneity for the co-moving inertial reference frame at  $p$ . Thus, in relativistic mechanics, one requires that  $\eta_{ab}\zeta^a F^b = 0$  (i.e. the four-velocity and the four-force are orthogonal; since the four-velocity is always a timelike vector, the four-force is always spacelike) and therefore,<sup>7</sup>

$$\eta_{ab}\zeta^a(\zeta^c\nabla_c\zeta^b) = 0. \quad (9)$$

Equation (9) together with the compatibility condition between metric and the derivative operator,  $\eta_{ab}\nabla_c = 0$ , again places a restriction on the *arbitrary* parameter  $s$  that can be used in the relativistic law of motion, equation (8). Again, Friedman shows that  $s$  must be linearly related to  $\tau$ .

This restriction on what the parameter  $s$  can be in the relativistic law of motion ‘has important implications’ according to Friedman. In the Newtonian context, the  $x_0$ -coordinate in an arbitrary inertial system is a suitable time function  $t$  because it is always linearly related to the scalar that parametrises the curve  $\sigma$ . So in any inertial coordinate system, equation (7) becomes the familiar expression of Newton’s second law:

$$m\frac{d^2x^i}{dt^2} = F^i. \quad (10)$$

Friedman then points out that this is not true in the relativistic case, because

the coordinate time  $t = x_0$  of an arbitrary inertial coordinate system is linearly related to  $\tau$  just in case the curve  $\sigma(\tau)$  in question has *constant* velocity [...]. So  $t$  is a suitable parameter in (29) [our equation (8)] only if  $\sigma(\tau)$  is a geodesic (and therefore  $F = 0$ ) (Friedman, 1983, p. 144).

He then explains this in yet another way. For inertial motions, we can regard the curve  $\sigma$ , whose tangent vector field is  $\zeta^c$ , as parametrised *either* by  $t$  or  $\tau$ . Although Friedman does not say this explicitly, this is because if  $\sigma$  is a timelike geodesic, i.e. the trajectory of a particle moving inertially, then the geodesic equation  $\zeta^c\nabla_c\zeta^b = 0$  will continue to hold even if we reparametrise the curve by *any* parameter  $s$  which is linearly related to  $\tau$ .<sup>8</sup> The parameter  $t$  satisfies this condition only when the curve is a geodesic (and hence  $F^b = 0$ ), so we can interchange the parameters. For example, consider a particle that moves with constant three-velocity  $u$  relative to some inertial frame  $F$ . Let its worldline be the image of the curve  $\sigma$  parametrised by the proper time  $\tau$ . The tangent field to

<sup>7</sup> Recall from the definition of  $F^b$  that  $\zeta^c\nabla_c\zeta^b$  is a vector parallel to  $F^b$ .

<sup>8</sup> One can prove the following general result. Let  $M$  be a manifold with a derivative operator  $\nabla_c$ . Let  $\gamma: I \subseteq \mathbb{R} \rightarrow M$  be a smooth curve with tangent field  $\zeta^b$ . Let  $\underline{\gamma} = \gamma \circ \alpha$  be a reparametrization of  $\gamma$ , i.e.  $s = \alpha(t)$  and  $\alpha$  is a diffeomorphism  $\alpha: I \rightarrow I$ . Then  $\underline{\gamma}$  is a geodesic if and only if  $s$  is linear in  $\tau$ , i.e.  $d^2s/d\tau^2 = 0$ .

this curve satisfies the geodesic equation. Now let  $F'$  be another inertial frame in ‘standard configuration’ which moves with constant velocity  $\mathbf{v}$  relative to  $F$ . If we parametrise the curve  $\sigma$  by the  $x_0$ -coordinate  $t$  of the coordinate system adapted to  $F'$ , then the tangent field of this curve will also satisfy the geodesic equation, since

$$\frac{d\tau}{dt} = \sqrt{1 - v^2} \quad (11)$$

implies

$$\frac{d^2\tau}{dt^2} = 0. \quad (12)$$

In other words,

$$\tau = \frac{1}{\gamma}t + k, \quad (13)$$

where  $\gamma$  is the Lorentz factor and  $k$  is an arbitrary constant. This is, of course, as it should be; for it tells us that any two observers will agree that the trajectory traced by an inertially moving particle is a geodesic path of Minkowski space-time regardless of how they parametrise the curve.

For non-inertial motions,  $t$  and  $\tau$  are not interchangeable because when the curve is not a geodesic,  $t$  is no longer linearly related to  $\tau$ . In other words, when we are considering the motion of particles subject to a force (i.e. particles that follow non-inertial paths for which  $F^b \neq 0$ ), we cannot simply interchange the parameters  $t$  and  $\tau$ . In this case, how  $\tau$  is related to  $t$  will change from ‘instant to instant’ as one ‘moves along the curve’. Again, physically this makes sense because it implies that any two observers will always agree that the path of a body moving non-inertially is not a geodesic. However, they will disagree on the magnitude of the particle’s four-acceleration *unless* they agree to parametrise the curve by the proper time  $\tau$ . This is because when  $\zeta^c \nabla_c \zeta^b = 0$ , the magnitude of this vector (at same point  $p$  on the image of the curve) will change depending on the parameter (see, for example, Schutz, 1985, p. 129). So in order to have a well-defined notion of four-force, one must always parametrise the curve by proper time. Thus, we arrive, with Friedman, at the following result. *When considering non-inertial motion, the worldlines of particles must always be parametrised by their proper time.*

Friedman’s derivation of mass–energy equivalence begins with the observation that not being able to interchange the parameters  $t$  and  $\tau$  in the equation of motion (8) ‘is responsible for the characteristic differences between relativistic dynamics and Newtonian dynamics’ (p. 145). He then proceeds to introduce the definitions for the ‘new’ dynamical quantities. First, he notes that in an inertial coordinate system, equation (8) becomes:

$$m_0 \frac{d^2x^\mu}{d\tau^2} = F^\mu. \quad (14)$$

Using the definition of the four-velocity (what Friedman calls the ‘proper velocity’),  $\zeta^\mu = dx^\mu/d\tau$ , one can re-write this equation as,

$$\frac{d}{d\tau}(m_0 \zeta^\mu) = F^\mu. \quad (15)$$

From this equation (15), Friedman concludes that the four-vector  $p^\mu = m_0 \zeta^\mu$  is ‘therefore the relativistic counterpart of the classical momentum’. Now consider a particle that moves with a ‘coordinate three-velocity’  $\mathbf{u} = (u_1, u_2, u_3)$  relative to some inertial frame  $F$ . The particle’s four-velocity  $\zeta^a$  has components  $\zeta^\mu$  in this inertial frame which are given by:

$$(\zeta^\mu) = \gamma(u)(1, u_1, u_2, u_3). \quad (16)$$

The components of the four-momentum in this inertial frame are:

$$(p^\mu) = m_0 \gamma(u)(1, u_1, u_2, u_3). \quad (17)$$

Friedman now asserts that the quantity  $\mathbf{p} = m_0 \gamma(u) \mathbf{u}$  is the relativistic three-momentum, which has magnitude  $p = m_0 \gamma(u) u$ . He then defines the ‘relativistic mass’ of the particle by the relation  $m = p/u$  and thus arrives at the expression  $m = m_0 \gamma(u)$  which is ‘the traditional formula for the dependence of mass on velocity’ (p. 145).

Friedman now uses the following equation to define the relativistic energy of the particle:

$$\frac{dE}{dt} = u \frac{dp}{dt}, \quad (18)$$

where  $p$  is the magnitude of the relativistic momentum. There are two important things to notice about this definition. First, the parameter with respect to which we are differentiating is the coordinate time  $t$  and *not* the proper time  $\tau$ . In effect, this equation defines the relativistic energy of a particle relative to an inertial frame whose  $x_0$ -coordinate is  $t$ . Second, this definition is, again, motivated by the ‘corresponding’ definition from the Newtonian case. If we simply replace  $p$  by the magnitude of the *classical* momentum, then this equation can be used to define the classical energy of a particle relative to an inertial frame, viz solving (18) with  $p = mu$  yields

$$E = \frac{1}{2} mu^2 + k, \quad (19)$$

where  $k$  is a constant of integration. In the relativistic case, i.e. when we take  $p = m_0 \gamma(u) u$ , we find

$$E = m_0 \gamma(u) + k, \quad (20)$$

where  $k$ , again, is a constant of integration. Friedman then points out that it is usual practice to set  $E_0 = 0$  and ‘to define the relativistic energy’ by

$$E = m_0 \gamma(u) = m \quad (\text{with } c = 1). \quad (21)$$

He then claims that this is ‘the traditional relation between relativistic energy and relativistic mass’. Finally, he points out that if we compare equation (21) with equation (17), we ‘see’ that the four-momentum has components  $(E, \mathbf{p})$  where  $E$  is the relativistic energy and  $\mathbf{p}$  is the relativistic three-momentum: ‘This is why  $p^i$  [i.e. the four-momentum] is called the *energy–momentum* vector.’

This demonstration of mass–energy equivalence is not nearly as cavalier as the one Einstein considers in the opening of his (1935) paper, though it still fails in similar respects. Friedman has clearly shown how the change in the structure of spacetime leads to equations like (15) which suggest what the relativistic ‘counterparts’ of the classical dynamical quantities ought to be. In effect, Friedman has shown that if we assume, for example, that  $m\gamma$  can correctly be interpreted as the relativistic energy of a particle, then we can offer some justification for this interpretation. We can show how this quantity arises as the ‘relativistic counterpart’ of the familiar Newtonian quantity when the structure of spacetime changes. Still, Friedman has not shown the extent to which we are committed to these interpretations if we accept the physical principles that lie at the core of SR. Since Friedman relies on analogies to introduce the new relativistic quantities, Einstein’s criticisms are still pressing.

First, as I indicated in Section 3, Einstein is concerned that the quantities we routinely define as the relativistic momentum and relativistic energy may not satisfy the conservation conditions that we take to be definitive characteristics of these quantities. Friedman has not given any sort of argument to support the claim that the quantities that he defines satisfy this requirement. To illustrate this, consider two particles of equal rest-mass  $m_0$  travelling in opposite directions with equal speeds with respect to some inertial frame  $F$ . Suppose the particles experience a perfectly inelastic collision forming a doublet at rest at the origin of the coordinate system adapted to  $F$ . From what Friedman has demonstrated we can determine, given the relative state of motion of an observer, the value for the mass of each particle *before* the collision for this observer. Now suppose we ask the following question: what is the mass of the resulting doublet? To answer this question, Friedman must use (1) the quantity  $m_0\gamma$ , which he interprets as the energy of each particle, and (2) the principle of conservation of energy. Einstein’s objection is that Friedman cannot apply the principle of conservation of energy to the quantity  $m_0\gamma$  unless we have good reason to suppose that this quantity *should* be interpreted as the relativistic energy of the particle. But if all we have is an analogical argument, which is in the end all that Friedman offers when he defines the relativistic energy, we cannot accept this interpretation since we have no reason to suppose that this quantity will be conserved in all collisions. This is not circular. What Einstein is saying is that we must know that for all collisions the quantity  $m_0\gamma$  is conserved *before* we say that it is so conserved for *this* collision.

Second, Einstein (1935) criticises the purported demonstration he analyses because it does not offer a proof of ME1. Friedman’s demonstration also fails to

provide an argument to support this claim. This, it seems to me, is a result of being concerned primarily with changes to the dynamical quantities that arise as a result of changes to the velocity of a particle. In other words, Friedman does not consider the ‘other changes’, which are not merely changes to the translational velocities of particles, that give rise to ME1. As a result, Friedman’s demonstration does not clearly show the important difference between the classical concept of inertial mass and the relativistic concept of rest-mass. As Torretti (1990, pp. 58–68) has pointed out, the key difference between these two concepts is best understood by looking at their associated structure. For Torretti, this means, roughly, that while the classical concept of inertial mass can be understood as an extensive magnitude that behaves like an element of the additive semi-group of the real numbers, the relativistic rest-mass does not. This is because while instances of the classical concept of mass combine in a way that can be modelled by the addition of real numbers, the rest-mass of relativistic mechanics does not. As Mermin and Feigenbaum explain, ‘if a particle in its rest frame decays into a collection of particles with a total kinetic energy  $K$ , then the total mass of the final group of particles must be less than the initial mass by  $K/c^2$ , (1990, p. 270).

In sum, Friedman offers some justification for the introduction of the ‘new’ relativistic dynamic quantities when he considers how changes to the structure of spacetime affect the classical definitions. Thus, he has given some justification for ME2 and, implicitly at least, he has shown that ME3 is not really the significant aspect of the equivalence of mass and energy because one can always change the zero-point of the energy scale. However, Friedman seems to assume the interpretation of the quantities  $m\gamma$ ,  $m\gamma\mathbf{u}$ ,  $m(\gamma - 1)$ . He has not clearly shown the extent to which we are committed to the interpretations of the quantities he defines as relativistic momentum and kinetic energy. For example, given Friedman’s presentation, if experiments revealed that the quantity he defines as the *relativistic* energy was not conserved, one seems free to postulate another definition of relativistic energy as long as one does not use Newtonian absolute time in the definition. Of course, this would be suspect methodology and one not likely to withstand critical scrutiny. My point is just that Friedman’s discussion does not give us a clear sense of all the physical principles and postulates of SR that would be at stake if the equivalence of mass–energy were not empirically corroborated. Finally, as I have argued above, Friedman does not offer a proof of ME1, which is really the significant result associated with Einstein’s famous equation.

## 5. Einstein’s 1935 Dynamical Derivation of $E=mc^2$

Einstein’s (1935) derivation of mass–energy equivalence offers stronger motivations for interpreting the quantities  $m\gamma$ ,  $m\gamma\mathbf{u}$ , and  $m(\gamma - 1)$  as the relativistic energy, momentum, and kinetic energy, respectively, than we have seen in Friedman’s derivation. Einstein’s explicit aim in this paper is to

show that:

If the principles of conservation of impulse and energy are to hold for all coordinate systems which are connected with one another by the Lorentz transformations, then impulse and energy are really given by the above expressions and the presumed equivalence of mass and rest-energy also exists (p. 225).

The antecedent of the conditional deserves some comment. Einstein assumes the principle of conservation of momentum and the principle of conservation of energy along with the two postulates of SR (which lead to the Lorentz transformations). What is noteworthy is that although Einstein assumes these conservation principles and the *kinematical* framework of SR, he does not assume the exact mathematical expressions that define the ‘new’ dynamical quantities but only assumes their general ‘form’. Einstein does not assume, in other words, that, for example,  $m\gamma$  should be interpreted as the relativistic energy of a particle of mass  $m$ . Instead, he assumes two general physical principles, which we have no reason to suppose no longer hold despite the change in the structure of spacetime, and other ‘reasonable’ assumptions, like the fact that the energy of a particle should be a monotonically increasing function of the speed of the particle (assuming that space is isotropic). He then shows how these *theoretical* constraints, combined with the *methodological* constraint that the relativistic quantities should reduce to their Newtonian counterparts in the appropriate limit, single out the expressions for the ‘new’ dynamical quantities. He does this by showing that the only quantities that are conserved during a perfectly elastic collision between two particles are  $m\gamma\mathbf{u}$  and  $m(\gamma - 1)$ . Thus, while Einstein has not done the impossible—he has not shown that given the definitions of relativistic momentum and relativistic kinetic energy one can *prove* that these quantities are conserved without appealing to a conservation principle—he has given far stronger motivations for the interpretations of these quantities than one finds in derivations that proceed by analogy. In this way, Einstein establishes ME2 and ME3. He then *proves* ME1 by analysing an inelastic collision. Since Einstein’s (1935) derivation is not well known to philosophers, I proceed through the derivation in some detail, picking up where we left off in Section 2.

Einstein assumes that the relativistic momentum and relativistic energy for a particle of mass  $m$  have the following form:

$$p_\mu = mu_\mu F(u), \quad \mu = 1, 2, 3 \quad \text{and} \quad E = E_0 + mG(u), \quad (22)$$

where  $E_0$  is the rest-energy of the particle and  $p_\mu$  are the components of the relativistic momentum in some inertial coordinate system  $S$ , and  $F$  and  $G$  are arbitrary (continuous) functions to be determined. The expression for the relativistic kinetic energy then has the form:

$$KE_{\text{rel}} = mG(u). \quad (23)$$

These expressions for the relativistic momentum and relativistic kinetic energy are suggestive, *prima facie*, because according to Taylor’s theorem, they have the

potential to reduce to their Newtonian counterparts in the appropriate limit.<sup>9</sup> The requirement that these definitions actually reduce to their Newtonian counterparts—a *methodological* constraint—imposes additional constraints on the functions  $F$  and  $G$ . For example, using the Taylor expansion for  $F$  yields the following expression for the relativistic momentum:

$$\mathbf{p}_{\text{rel}} = m\mathbf{u}[(F(0) + F'(0)u + \frac{1}{2}F''(0)u^2 + O(u^3))]. \quad (24)$$

This expression will reduce to its corresponding Newtonian counterpart if terms of the second order in  $u$  and higher are neglected and if  $F(0) = 1$  and  $F'(0) = 0$ . Similarly, the expression for  $KE_{\text{rel}}$  will reduce to its Newtonian counterpart if  $G(0) = 0$ ,  $G'(0) = 0$ ,  $G''(0) = 1$ , and if terms of the third order and higher in  $u$  are neglected.

Einstein further assumes that the ‘impulse and energy of the mass point are independent of the direction of motion and of the orientation of the mass-point relative to the velocity’ (p. 227). Thus, the functions  $F(u)$  and  $G(u)$  can be treated as functions of the speed of the particle, as I have written them. This assumption is warranted by our belief in the isotropy of space. Einstein also assumes that the  $F(u)$  and  $G(u)$  are ‘universal, even functions of the velocity  $u$ , which vanish for  $u = 0$ ’ (p. 226).<sup>10</sup> Finally, Einstein implicitly assumes that the functions  $F$  and  $G$  are monotonic bijections. This ensures that for every velocity there is a distinct value for the momentum (or kinetic energy) and that as the velocity increases, the momentum (or kinetic energy) also increases, i.e. we never find a lower value of the momentum (or kinetic energy) for a higher value of the velocity.

Given that one cannot prove the conservation principles by assuming only the definitions of the new dynamical quantities, Einstein sets out to prove the converse: *if* the conservation principles hold, regardless of the exact mathematical expressions of these quantities, then the definitions of relativistic momentum and relativistic kinetic energy must be those given by equations (4) and (5). Einstein’s goal is to determine the exact mathematical expressions for the functions  $F(u)$  and  $G(u)$  from the assumptions I have outlined above. What is distinctive about this approach is that Einstein is using the ‘principles’ of SR (including the conservation principles) to determine what the expressions for the ‘new’ relativistic dynamical quantities must be.

To find the expressions for the functions  $F(u)$  and  $G(u)$ , Einstein analyses an elastic, eccentric collision. He begins by considering a pair of particles of equal (inertial) mass  $m$  that travel with equal and opposite velocities toward each other when referred to a coordinate system  $S'$  (Einstein refers to such a pair of particles as a *particle pair*); their velocities are parallel (though of course

<sup>9</sup> Assuming, of course,  $F(u)$  and  $G(u)$  are continuous and have continuous derivatives.

<sup>10</sup> Since we are assuming that these are functions of the speed, we can take them as being defined on  $[0, \infty)$  instead of treating them as even functions of the velocity. Also, the requirement that both functions vanish for  $u = 0$  is not, strictly speaking, correct. As I have shown above, for the function  $F$  the requirement is that  $F(0) = 1$  though, of course,  $\mathbf{p}(0) = 0$ .

opposite in sign) but not collinear (since the collision is eccentric). To keep track of the particles, Einstein uses ‘+’ and ‘-’ signs. So, for example, we have  $\mathbf{u}'_+ = -\mathbf{u}'_-$ . Einstein then asks: with respect to  $S'$ , what are the velocities of the particles *after* the collision?

If the velocities after the collision were not likewise equal and opposite, the impulse law would be violated. If the common velocity of both masses after the collision were not equal in magnitude to that before the collision, in case the collision is elastic, the energy law would be violated. *This holds independently of the particular law of dependence of impulse and energy on the velocity* (1935, p. 227, emphasis added).

Let me separate the two claims contained in this answer. The first is that if the principle of conservation of momentum holds, then the velocities of both particles after the collision will also be equal and opposite, i.e.  $\underline{u}'_+ = \underline{u}'_-$ , where the barred quantities indicate value *after* the collision. Einstein’s claim is that this is true quite independently of what the exact mathematical expression for the function  $F(u)$  in the definition of the relativistic momentum turns out to be. Although Einstein does not provide an argument to support this claim, one can show that this is indeed the case. We begin with the law of conservation of momentum:

$$\Sigma \mathbf{p}_{\text{in}} = \Sigma \mathbf{p}_{\text{out}}. \quad (25)$$

For the *particle pair* in Einstein’s thought experiment we know that, since  $\mathbf{u}_+ = -\mathbf{u}_-$  before the collision, we have:

$$\Sigma \mathbf{p}_{\text{in}} = m\mathbf{u}_+F(u_+) + m\mathbf{u}_-F(u_-) = 0. \quad (26)$$

Therefore, by the law of conservation of momentum,

$$\Sigma \mathbf{p}_{\text{out}} = m\underline{\mathbf{u}}_+F(\underline{u}_+) + m\underline{\mathbf{u}}_-F(\underline{u}_-) = 0. \quad (27)$$

What we want to prove is that if  $\underline{u}_+ \neq \underline{u}_-$ , then  $\Sigma \mathbf{p}_{\text{out}} \neq 0$ . We show the contrapositive. From (27) we have:

$$\underline{\mathbf{u}}_+F(\underline{u}_+) + \underline{\mathbf{u}}_-F(\underline{u}_-) = 0, \quad (28)$$

that is,

$$\underline{\mathbf{u}}_+F(\underline{u}_+) = -\underline{\mathbf{u}}_-F(\underline{u}_-), \quad (29)$$

or

$$\underline{u}_+F(\underline{u}_+) = \underline{u}_-F(\underline{u}_-), \quad (30)$$

(taking the magnitude of both sides). Thus

$$\frac{\underline{u}_+}{\underline{u}_-} = \frac{F(\underline{u}_-)}{F(\underline{u}_+)}. \quad (31)$$

Equation (31) is satisfied if and only if  $\underline{u}_+ = \underline{u}_-$ , since  $F$  is a monotonically increasing function.<sup>11</sup> Therefore, if the conservation of momentum holds, regardless of the exact expression for the momentum,  $\underline{u}_+ = \underline{u}_-$ , i.e. the speeds of the particles are equal after the collision.

Einstein's second claim in the quotation is that if the common speed of both particles  $\underline{u} = \underline{u}_+ = \underline{u}_-$  after the collision is not equal to their common speed  $u$  before the collision, i.e. if  $u \neq \underline{u}$ , in the case when the collision is elastic, then the conservation of energy would be violated. Again, one easily shows the contrapositive. First, recall that the energy of a particle is given by  $E = E_0 + mG(u)$ . In the case of an elastic collision, the rest-mass and rest-energy of each particle before and after the collision remain the same. Therefore, energy conservation entails that for the *particle pair* consisting of two particles of mass  $m$  moving the common speed  $u$  in opposite directions, we have:

$$2E_0 + 2mG(u) = 2E_0 + 2mG(\underline{u}), \quad (32)$$

thus

$$G(u) = G(\underline{u}). \quad (33)$$

Again, since we assume  $G$  is a monotonically increasing function, equation (33) can only be satisfied if  $u = \underline{u}$ . Therefore, if energy conservation is satisfied, independently of the exact mathematical expression for the relativistic kinetic energy, the common speed of the *particle pair* before the collision is equal to the common speed of the *particle pair* after the collision. Putting these two results together leads to the conclusion that if two particles form a *particle pair* before a collision, they will also be a *particle pair* after the collision. The only difference is that the direction of their velocities will be different.

At this stage, Einstein has already derived the following equations for the components of the velocities of a *particle pair* from the Lorentz transformations:

$$\gamma(u_+) + \gamma(u_-) = 2\gamma(u')\gamma(v), \quad (34)$$

$$u_1 + \gamma(u_+) + u_1 - \gamma(u_-) = 2v\gamma(u')\gamma(v), \quad (35)$$

$$u_2 + \gamma(u_+) + u_2 - \gamma(u_-) = 0, \quad (36)$$

$$u_3 + \gamma(u_+) + u_3 - \gamma(u_-) = 0. \quad (37)$$

In these equations, the unprimed quantities are those measured in a coordinate system  $S$  which moves with velocity  $v$  relative to  $S'$ . Now since the *particle pair* is

<sup>11</sup> To see this, one considers three cases. *Case I:*  $u_+ > u_-$  and thus  $F(u_+) > F(u_-)$ , since  $F$  is monotonically increasing. However, in this case, equation (31) cannot hold since  $u_+/u_- > 1$  but  $F(u_+)/F(u_-) < 1$ .

*Case II:*  $u_+ < u_-$  and thus  $F(u_+) < F(u_-)$ , since  $F$  is monotonically increasing. Again, in this case, equation (31) cannot hold since  $u_+/u_- < 1$  but  $F(u_+)/F(u_-) > 1$ .

*Case III:*  $u_+ = u_-$  and  $F(u_+) = F(u_-)$ , and equation (31) satisfied.

turned into another *particle pair* with speed  $u'$  relative to  $S'$  after the collision, the right-hand side of equations (34)–(37) does not change after the collision. The left-hand side changes because all the velocities must be replaced by the ‘barred’ velocities. So we get another four equations like (34)–(37). After a little bit of algebra, one arrives at the following equations, which Einstein refers to as ‘conservation equations’ (1935, p. 227):

$$\gamma(u_+) + \gamma(u_-) = \gamma(\underline{u}_+) + \gamma(\underline{u}_-), \quad (38)$$

$$u_{i+}\gamma(u_+) + u_{i-}\gamma(u_-) = u_{i+}\gamma(\underline{u}_+) + u_{i-}\gamma(\underline{u}_-). \quad (39)$$

How do these equations help us find the functions  $F(u)$  and  $G(u)$ ? Einstein says, with characteristic brevity:

These equations, which are valid in general for elastic collisions of equal masses, have the form of conservation equations; it may therefore be taken for granted that no other symmetrical or anti-symmetrical functions of the velocity-components exist which in the present case of the elastic collision of two identically constituted material points give anything analogous (1935, p. 227).

Einstein is boldly stating that these conservation equations are unique. To understand this claim, recall how he arrives at these conservation equations. Einstein uses only the Lorentz transformations (and a little algebra) to arrive at equations (34)–(37). And since we are assuming that the Lorentz transformations are unique, i.e. they are the only allowable transformations between two inertial coordinate systems, equations (34)–(37) are also unique (these are the only transformation equations for the velocities of the *particle pair*). From the analysis of the elastic collision, we arrive at the result a *particle pair* before the collision is converted into another *particle pair* after the collision. The only thing that changes is the direction of the velocities. This, together with equations (34)–(37) when written for quantities before and after the collision, yields the conservation equations. But since equations (34)–(37) are unique, so are the conservation equations. In other words, there are only two ways in which we could obtain a different set of conservation equations. First, the equations would be different if the result of the thought experiment regarding the common speed of the particles were different. But this would only be the case if, as we have seen, the principle of conservation of energy or the principle of conservation of momentum were violated. Second, the equations would be different if equations (34)–(37) were different. But equations (34)–(37) would be different only if we accepted a set of transformation equations between inertial frames that was *not* the Lorentz transformations. Hence, equations (38) and (39) are the only equations of this type that obtain between the components of the velocities involved. Thus, given our assumptions about the uniqueness of the Lorentz transformations and the applicability of the conservation principles in analysing the collision, the conservation equations (38) and (39) are unique.<sup>12</sup>

<sup>12</sup> Strictly speaking, these equations are unique only up to an additive constant.

Einstein now states his conclusion about the functions  $F(u)$  and  $G(u)$ . He tells us, directly after the passage I have quoted above, that ‘consequently we shall have to regard’ the relativistic momentum and kinetic energy as given by the expressions:

$$\mathbf{p}_{\text{rel}} = m\mathbf{u}\gamma(u), \quad \text{and} \quad KE_{\text{rel}} = m(\gamma(u) - 1). \quad (40)$$

Thus, the conservation equations suggest that:

$$F(u) = \gamma(u), \quad \text{and} \quad G(u) = \gamma(u) - 1. \quad (41)$$

This last inference also requires some explanation. Since the collision we are considering is *elastic*, it follows that the masses of the particles, with respect to the inertial coordinate system  $S$ , before and after the collision, are the same. Consequently, we can write:

$$m\gamma_+ + m\gamma_- = m\underline{\gamma}_+ + m\underline{\gamma}_-, \quad (42)$$

$$m\mathbf{u}_{i+}\gamma_+ + m\mathbf{u}_{i-}\gamma_- = m\underline{\mathbf{u}}_{i+}\underline{\gamma}_+ + m\underline{\mathbf{u}}_{i-}\underline{\gamma}_-, \quad (43)$$

where  $\gamma_+ := \gamma(u_+)$  and so on. Looking back at the general mathematical form of the expressions for relativistic momentum and relativistic kinetic energy of equations (22) and (23), equations (42) and (43) ‘suggest’ that we take  $F(u) = \gamma(u)$  and  $G(u) = \gamma(u)$ . The disparity between this conclusion and Einstein’s is removed, perhaps not surprisingly, by appeal to the Newtonian limit. Quite simply, if we had taken  $G(u) = \gamma(u)$ , then the expression for the relativistic kinetic energy would *not* reduce to its classical counterpart in the Newtonian limit since there would be an extra additive constant  $m$ . Alternatively, we could note, as Einstein does, that taking  $G(u) = \gamma(u)$  does not satisfy the requirement that the kinetic energy vanish when the velocity is zero while  $G(u) = \gamma(u) - 1$  does. The latter expression will, of course, also be conserved in equation (42).

Once Einstein motivates the definitions for the relativistic momentum and kinetic energy, the derivation of mass–energy equivalence, ME1, is straightforward. Einstein now considers an *inelastic* collision and appeals to the principle of conservation of energy to arrive at the desired result. Consider again a *particle pair* consisting of two particles with equal rest-energies  $E_0$  that move with equal and opposite velocities with respect to  $S'$ . Einstein assumes that whatever internal changes the particles suffer as a result of the collision are equal to each other. Since the collision is *inelastic*, both  $E_0$  and  $m$  can change as a result of the collision. As we have seen, the conservation of momentum implies that the velocities of the particles after the collision must be equal and opposite. Assuming that the conservation of energy holds both in the inertial frame  $S'$  and in another frame  $S$  related to it by a Lorentz transformation, one can write down two conservation equations:

$$2E_0 + 2m(\gamma' - 1) = 2\underline{E}_0 + 2m(\gamma' - 1), \quad (44)$$

$$2E_0 + m(\gamma_+ - 1) + m(\gamma_- - 1) = 2\underline{E}_0 + \underline{m}(\underline{\gamma}_+ - 1) + \underline{m}(\underline{\gamma}_- - 1). \quad (45)$$

Using equation (34) one can rewrite these equations and subtract one from the other to obtain:

$$\underline{E}_0 - E_0 = \underline{m} - m. \quad (46)$$

From this equation, Einstein concludes: ‘The rest-energy changes, therefore, in an inelastic collision (additively) like the mass’ (1935, p. 229). Einstein adds that since the rest-energy is determined only up to an additive constant, ‘one can stipulate that  $E_0$  should vanish together with  $m$ ’ (1935, p. 229). If the rest-energy vanishes with  $m$ , which is clearly only a *hypothesis* here (and in all derivations of mass–energy equivalence in SR), then it follows that  $E_0 = m$ , ‘which states the principle of the equivalence of mass and rest-energy’ (1935, p. 229). Thus, Einstein *proves* ME1.

In this way, Einstein’s concern, that if we simply *define* the relativistic momentum and kinetic energy as in equation (40) the conservation laws might not hold, is quelled by showing that if the conservation laws hold, these expressions are uniquely singled out from a range of possible options. We call the quantities defined by these equations ‘relativistic momentum’ and ‘relativistic kinetic energy’ not only because they reduce to their classical counterparts but because they obey the same conservation principles—relativistic momentum and relativistic kinetic energy are conserved in elastic collisions. Thus, Einstein has shown that subject to a collection of physical assumptions, the mathematical expressions for these particular physical quantities are uniquely determined. From these definitions we get ME2; Einstein then gives a proof of ME1.

## 6. Conclusion

Einstein’s (1935) derivation of mass–energy equivalence is philosophically significant, because it shows that adopting the traditional definitions for the ‘new’ relativistic dynamical quantities is motivated not merely by analogy but by something far deeper, viz the physical principles that lie at the core of the theory. Thus, Einstein’s (1935) derivation, and other derivations of this kind, show that much more than an arbitrary definition or analogy is at stake as we wait for the experimental data to verify these results. If, contrary to fact, the equivalence of mass–energy were not empirically corroborated, all the physical principles that motivated the definitions in the first place would be impugned. Finally, Einstein’s (1935) derivation, and others that proceed along these lines, enable us to separate three distinct physical consequences of the equivalence of mass–energy that are not always clearly separated and to see what aspects of the definitions of the new dynamical quantities are conventional.

*Acknowledgement*—A portion of the research for this paper was made possible by the generous support of the Social Sciences and Humanities Research Council of Canada and The University of Western Ontario. I would also like to thank Professors Robert DiSalle, Rob Clifton and William Demopoulos and two anonymous referees for their valuable comments.

## References

- Einstein, A. (1905a) 'On the Electrodynamics of Moving Bodies', in Lorentz *et al.* (1952), pp. 35–65.
- Einstein, A. (1905b) 'Does the Inertia of a Body Depend Upon Its Energy-Content?', in Lorentz *et al.* (1952), pp. 69–71.
- Einstein, A. (1935) 'Elementary Derivation of the Equivalence of Mass and Energy', *American Mathematical Society Bulletin* **41**, 223–230.
- Einstein, A. (1950) 'On the Generalized Theory of Gravitation', in *Ideas and Opinions*, translated by S. Barmann (New York: Bonanza Books) pp. 341–356.
- Feynman, R. P., Leighton, R. B. and Sands, M. L. (1963) *The Feynman Lectures on Physics*, Vol. 1 (Reading, MASS: Addison-Wesley).
- Friedman, M. (1983) *Foundations of Spacetime Theories* (Princeton, NJ: Princeton University Press).
- Lewis, G. N. and Tolman, R. C. (1909) 'The Principle of Relativity, and Non-Newtonian Mechanics', *Philosophical Magazine* **6**(18), 510–523.
- Lorentz, H. A., Einstein, A., Minkowski, H. and Weyl, H. (1952) *The Principle of Relativity* (New York: Dover).
- Mermin, D. and Feigenbaum, J. (1990) ' $E = mc^2$ ', in D. Mermin, *Boojums All the Way Through* (Cambridge: Cambridge University Press).
- Schutz, B. (1985) *A First Course in General Relativity* (Cambridge: Cambridge University Press).
- Stachel, J. and Torretti, R. (1982) 'Einstein's First Derivation of Mass–Energy Equivalence', *American Journal of Physics* **50**(8), 760–761.
- Torretti, R. (1990) *Creative Understanding: Philosophical Reflections on Physics* (Chicago: The University of Chicago Press).