Interpretations of Einstein's Equation $E = mc^2$

Francisco Flores

Interpretations of Einstein's equation differ primarily concerning whether $E = mc^2$ *entails that mass and energy are the same property of physical systems, and hence whether there is any sense in which mass is ever 'converted' into energy (or vice versa). In this paper, I examine six interpretations of Einstein's equation and argue that all but one fail to satisfy a minimal set of conditions that all interpretations of physical theories ought to satisfy. I argue that we should prefer the interpretation of Einstein's equation that holds that mass and energy are distinct properties of physical systems. This interpretation also carries along the view that while most cases of 'conversion' are not genuine examples of mass being 'converted' into energy (or vice versa), it is possible that the there are such 'conversions' in the sense that a certain amount of energy 'appears' and an equivalent of mass 'disappears'. Finally, I suggest that the interpretation I defend is the only one that does not blur the distinction between what Einstein called 'principle' and 'constructive' theories. This is philosophically significant because it emphasizes that explanations of Einstein's equation and the 'conversion' of mass and energy must be top-down explanations.*

1. Introduction

Einstein's equation $E = mc^2$ has received two main types of interpretations. First, some philosophers and physicists have suggested that Einstein's equation tells us whether the *properties* mass and energy are the same. Second, some philosophers and physicists have gone further to claim that Einstein's equation has ontological consequences concerning the fundamental stuff of modern physics. Each interpretation of Einstein's equation includes, either explicitly or implicitly, a view concerning how we ought to understand purported 'conversions' of mass and energy. So, for example, Bondi and Spurgin (1987) have argued that while we are not entitled to infer from Einstein's

Francisco Flores is at California Polytechnic State University, USA
Correspondence to: Francisco Flores, Department of Philosophy, California Polytechnic State University, 1 Grand Avenue, San Luis Obispo, CA 93407-0327, USA.

equation that mass and energy are the *same* property, there is no sense in which mass is ever 'converted' into energy (or vice versa).

My purpose here is to argue for the interpretation of Einstein's equation that holds that mass and energy are distinct properties of physical systems and that it is possible that there are cases where mass is *converted* into energy in the sense that a certain amount of mass 'disappears', and an equivalent amount of energy 'appears' (or vice versa). I reach this conclusion by first showing that all of the leading and influential interpretations of Einstein's equation in the literature (which I describe in Section 2), except the one I favor, fail to satisfy three minimal constraints on interpretations of physical theories. The first requirement I impose is simply the familiar one that that an interpretation I of a theory T should tell us what the world is like if T is true. I then impose the two additional requirements that I should not appeal to hypotheses *outside* T, and that I should interpret the mathematical formalism of T uniformly (see Section 3). I argue that Lange's (2001, 2002) recent interpretation of Einstein's equation fails to satisfy the third constraint. The other interpretations I consider, except the one I favor, fail to satisfy the second constraint because they assume hypotheses concerning the nature of matter that lie *outside* special relativity (Section 4). Finally, I suggest that the latter failure is significant because it involves us in the quest for bottom-up type explanations of $E = mc^2$ and the 'conversion' of mass and energy. However, as I have argued elsewhere (Flores 1999), special relativity is a principle theory, and principle theories only offer top-down explanations (Section 5). I begin (in Section 1) with some preliminary remarks outlining my assumptions for this paper.

2. Preliminaries

My main assumptions throughout this paper concern the notions of mass in Newtonian and relativistic physics. I will assume that in Newtonian physics, one can define *inertial mass* as a measure of how a body responds to changes in velocity and that inertial mass can be measured dynamically along the lines first suggested by Maxwell and Mach. In the context of special relativity, I will assume that the rest-mass of a body is a measure of its inertia. Thus, I will assume that rest-mass is the closest analogue in relativistic physics of Newtonian inertial mass despite important differences concerning how instances of rest-mass combine arithmetically.

The philosophical significance of these assumptions is that I shall not discuss interpretations of Einstein's equation that claim that *matter* is convertible into energy (or vice versa). Such interpretations, though common in popular expositions of Einstein's equation, rest at best on a mistaken adherence to Newton's notion of mass as a measure of the quantity of matter, which Maxwell, Mach, and others correctly criticized. At worst, such interpretations either assume that energy is a type of substance, which is also untenable for well-known reasons, or make a type of category mistake. Einstein's equation does not tell us that a substance, viz., *matter*, is transformed into a property, viz., energy.

I will also assume the familiar distinction between relativistic mass and rest-mass, though I will focus exclusively on the latter especially because the most philosophically interesting and challenging consequences of Einstein's equation concern the equality (in units in which $c = 1$) of rest-energy and rest-mass. I will follow the common practice of dropping the adjective 'rest' from 'rest-mass' unless ambiguity threatens. From now on, then, I will use the equation $E = mc^2$ to designate the numerical equivalence of rest-energy and rest-mass, which I will simply call 'mass–energy equivalence'. Thus, I will not use Einstein's equation to designate, as it usually does, the numerical equivalence of the total relativistic energy and relativistic mass.

Finally, I will assume the familiar 'geometric' or 'co-variant' formulation of special relativity. On this view, the mass of a body is represented within the mathematical formalism as the magnitude of that body's four-momentum. The rest-energy, as it is called in presentations of relativity that emphasize coordinate transformations, is then simply a scalar multiple of a scalar invariant.

3. Interpretations of Einstein's Equation

There are six interpretations of Einstein's equation I wish to consider. All these interpretations treat mass and energy as properties of physical systems. Four of these interpretations, which I will call *property* interpretations, confine their claims to the question of whether mass and energy are the same property, and whether there is a concrete sense in which mass and energy are 'convertible'. The other two interpretations I consider, which I will call *ontological* interpretations, take the additional step of drawing a further conclusion concerning the fundamental stuff of modern physics from mass–energy equivalence. The two types of interpretations are closely related. As I shall presently show, the ontological interpretations I consider are based on a particular type of property interpretation, which claims that mass and energy are the *same* property. My goal in this section is to describe each interpretation of Einstein's equation and to specify its stance concerning the 'conversion' of mass and energy.

The first *property* interpretation I wish to consider, which is advocated by physicists and philosophers such as Eddington (1929) and Torretti (1996), is that Einstein's equation entails that mass and energy are actually the same property. According to this interpretation, which I will call the *same-property* interpretation, mass and energy were hitherto regarded as distinct properties because in Newtonian physics they are, and must be, measured in different units. However, because, according to special relativity, light travels at the same speed for all inertial observers, one can select units such that spatial intervals are specified in units of time, e.g. light-years. When such units are chosen, energy and mass have the same units and are numerically equal. Hence, mass and energy are not two distinct properties after all. According to Torretti (1996), the only reason that we ever regarded mass and energy as distinct properties in the first place is that we do not perceive spatiotemporal intervals directly. Instead, we perceive spatial intervals and temporal intervals separately, and we perceive them differently. This leads us to associate different units to space and time, and hence to mass and energy.

Proponents of the *same-property* interpretation say little about how one ought to understand the purported 'conversion' of mass and energy as a physical process. The reason is clear. If mass and energy really are the same property, then there is no sense in which one can be *converted* into the other through some kind of physical transformation. At best, the 'conversion' of mass and energy is a conversion between two kinds of units akin to the conversion from meters to yards. Consequently, I will refer to this interpretation as the *same-property*, *no-conversion* interpretation of Einstein's equation to indicate that, according to this interpretation, mass is not physically transformed into energy (or vice versa). Instead, according to Torretti (1996), cases where there *appears* to be a conversion, are really cases where there is a change in the distribution of the one property, call it 'mass–energy', among the parts of a physical system. Thus, for example, when a body *appears* to radiate energy, special relativity teaches that it is really radiating mass–energy.

The second *property* interpretation of mass–energy equivalence stands in contrast to the *same-property* interpretation, for it holds that mass and energy are *different* properties of physical systems. There are two slightly different versions of this interpretation depending on what one says about the 'conversion' of mass and energy. According to the first version, proposed by Bondi and Spurgin (1987), there is no such thing as a 'conversion' of mass and energy. Thus, I will call Bondi and Spurgin's interpretation the *different-properties*, *no-conversion* interpretation of Einstein's equation. According to the second version, which is adopted, for example, by Rindler (1977), genuine conversions of mass and energy are possible according to Einstein's equation, and such conversions have been observed. I will call this interpretation the *different-properties, conversion* interpretation of mass–energy equivalence. I will treat these two versions of the *different-properties* interpretation separately and discuss them in turn.

Bondi and Spurgin (1987) have argued against the *same-property* interpretation in their criticism of the view that mass and energy are 'inter-convertible'. According to Bondi and Spurgin, Einstein's equation does not tell us that mass and energy are the same property any more than the equation $M = \rho V$ tells us that mass and volume are the same property. Bondi and Spurgin argue that mass and energy, like mass and volume, have different dimensions, and hence are different properties. For now, let me simply observe that Bondi and Spurgin's argument is about the *dimensions* of mass and energy, and not the *units* in which they are measured. I will return to this observation in Section 5.

Concerning the 'conversion' of mass and energy, Bondi and Spurgin argue that all cases of purported 'conversions' are really cases where energy of one kind is transformed into energy of another kind. Mass is never 'converted' into energy, or vice versa. For example, suppose we heat a billiard ball. As the ball absorbs heat energy, its mass will increase according to Einstein's equation. Thus, one might ordinarily say, this is an example of energy being converted into mass. Bondi and Spurgin argue that these examples are at best highly misleading. No additional mass is added to the billiard ball as it is heated. Instead, the physical process that is taking place is that heat energy from the source is being transformed into the kinetic energy of the microscopic constituents of the billiard ball. It is this increased energy of the microscopic constituents that *contributes* to the mass of the billiard ball. However, the heat energy never *becomes* mass: it is simply transformed into kinetic energy.

Bondi and Spurgin argue that the best way to interpret Einstein's equation is to say that 'energy has mass', which is an unfortunate locution. 'Energy has mass' is superficially similar to 'water has mass'. The latter, but not the former, is an attribution of a property to a substance. Bondi and Spurgin clearly treat energy as a property, not as a substance. However, if one treats energy as a property, then 'energy has mass' suggests that mass can be the property of a property, i.e. a second-order property. This is unfortunate because mass is clearly a first-order property when we talk, for example, about the mass of a particle. Presumably, Bondi and Spurgin have no desire to introduce some fine metaphysical distinction between mass as a first-order property and mass as a second-order property. It is much more likely that in their quest for a simple and succinct formulation of Einstein's equation, they have gone a bit too far. It might have been better for them to say 'energy *contributes* to mass' or 'the mass of an object depends upon its energy-content', to borrow liberally from Einstein (1905).

Proponents of the *different-properties, conversion* interpretation, such as Rindler (1977), seem to hold that *some* purported cases of the 'conversion' of mass and energy are just as proponents of the *different-properties, no-conversion* interpretation describe. For example, energy is not converted into mass when one heats a macroscopic object. In such cases, there is merely a transformation of energy. However, according to the *different-properties*, *conversion* interpretation, *if* we reach a 'fundamental' level of matter at which the constituents of matter are philosophical atoms, then at this level there is a 'conversion' of mass and energy in the sense that a certain amount of one 'disappears' and an equivalent amount of the other 'appears'. Of course, whether there is such a fundamental level, and hence this sort of 'conversion', is not a *consequence* of $E = mc^2$. Einstein's equation merely imposes the restriction that *if* a certain amount of mass (say) disappears from a physical system, then an equivalent amount of energy must appear in the same physical system. Thus, proponents of the *different-properties*, *conversion* interpretation hold that, for example, pair annihilation reactions are genuine cases of mass being converted into energy.

The fourth property interpretation of Einstein's equation I wish to consider also stands against the *same-property* interpretation. The only proponent of this interpretation of whom I am aware is Lange (2001, 2002). Like Bondi and Spurgin, Lange develops his interpretation as part of an argument against the *same-property* interpretation and against the view that mass can be 'converted' into energy. According to Lange, mass is a real property of physical systems, since it is Lorentz-invariant. Energy, on the other hand, is not a real property since it is not Lorentz-invariant. Consequently, mass and energy cannot be the same property or measure the same thing since only one of them is real.

The claim that mass is a real property but energy is not places Lange's interpretation in the *no-conversion* camp. For there can be no physically interesting sense of 'conversion' that can accommodate a change from a 'real' property to a 'non-real' one. Lange is clearly aware of this. He asks, 'in what sense can mass be *converted* into energy when mass and energy are not on a par in terms of their reality?' (Lange 2002, 227). Furthermore, Lange argues that the purported 'conversion' of mass and energy is an illusion that arises from a change in our perspective when we shift from analyzing a physical system at a microscopic level to analyzing the same system at a macroscopic level. His argument for this conclusion is similar to Bondi and Spurgin's. If we heat a sample of gas, the heat energy from the source is transformed into kinetic energy of the particles that constitute the gas. However, the heat energy does not *become* mass. The difference with Lange is that he stops short of saying that energy contributes to mass. Instead, he concludes that *we* convert the energy of the constituents of the gas into mass when we shift our perspective and treat the gas as a single body. I shall refer to Lange's interpretation as a *one-property, no-conversion* interpretation of Einstein's equation.

I have so far described four versions of the property interpretation of Einstein's equation: the *same-property*, *no-conversion* interpretation, the *different-properties*, *no-conversion* interpretation of Bondi and Spurgin, the *different-properties, conversion* interpretation of Rindler, and finally Lange's recent *one-property*, *no-conversion* interpretation. My goal has been only to describe these interpretations, though I have done this in something of a dialectical way to capture some of the core differences among the interpretations. I now wish to describe two *ontological* interpretations of Einstein's equation.

One need not think of mass as a measure of the quantity of matter and of energy as a substance in order to draw ontological conclusions from mass–energy equivalence. Two noteworthy attempts to draw such conclusions are that of Einstein and Infeld (1938) and Zahar (1989). These two interpretations are closely related, for both hold that since there is no distinction between mass and energy as properties, and since it is by these properties that we distinguish in classical physics between matter and fields, we can no longer distinguish matter and fields. However, the two interpretations differ on the ontological conclusion they draw from these observations.

Zahar's position seems to be that our inability to distinguish between matter and fields suggests that the fundamental stuff of modern physics is a certain 'I-know-not-what', which can manifest itself either as matter or as field. I say that this is what Zahar's position *seems* to be because Zahar's presentation is somewhat clouded by ambiguities in his use of the terms 'mass', 'matter', 'energy', and 'field'. For example, according to Zahar, Einstein showed 'that 'energy' and 'mass' *could* be treated as two names for the same basic entity' (p. 262). Here, Zahar seems to be treating the terms 'mass' and 'energy' as terms that designate *substances*. Zahar then goes on to say that Einstein's equation teaches us that what is real 'is no longer the familiar hard substance but a new entity which can be interchangeably called matter or energy' (p. 263). In this passage, Zahar now seems to be treating the term 'matter' as the term that designates a substance, and energy again is treated as a term that designates a substance. It is because of passages such as these that Lange has characterized Zahar's interpretation as claiming that mass–energy equivalence entails that one type of stuff, viz., matter, is converted into another type of stuff, viz., energy (Lange 2001, 221). However, a more charitable interpretation of Zahar suggests otherwise.

Let us suppose that Zahar is aware with the familiar arguments raised against him by Lange (2001) concerning mass and energy not being measures of amount of stuff in the context of special relativity. What then becomes of Zahar's interpretation? Zahar can clearly adopt the view that mass is a measure of inertia, i.e. of a body's resistance to changes in velocity, and hence a *property* of matter. Indeed, Zahar seems to do just this in his discussion of general relativity (pp. 270ff.). He can furthermore agree, as he seems to do when he talks about the kinetic energy of a particle, that energy is a *property* of physical systems. On this reading, Zahar would be committed to the view that mass and energy are properties, and not measures of amount of stuff. Consequently, Lange's arguments against him would miss the mark. Furthermore, Zahar could continue to hold that matter and fields are distinguished in classical physics because they bear different properties, viz., matter has both mass and energy, whereas fields only have energy. Zahar could then argue that since the distinction between mass and energy is erased by Einstein's equation, and since these are the only properties that distinguish matter and fields, it follows that there is no longer a distinction between matter and fields. He could then go on to postulate that we now have only one entity, an 'I-know-not-what' that is neither matter nor field, but which can manifest itself as either. Finally, even if Lange is correct in his characterization of Zahar's position, the version of Zahar's position I have just sketched seems to be a plausible way, prima facie at least, to draw an ontological conclusion from mass–energy equivalence.

The second ontological interpretation of Einstein's equation, due to Einstein and Infeld (1938), is only minimally different from the interpretation I have reconstructed on behalf of Zahar. According to Einstein and Infeld, since we can no longer distinguish between mass and energy as properties, we can no longer distinguish between matter and field. At times, Einstein and Infeld seem to leap to the conclusion that therefore it follows that the fundamental stuff of physics is fields. However, in other places, they are a bit more cautious and state that it is at least possible that one can construct a physics with only fields in its ontology. Thus, both of the ontological interpretations of Einstein's equation I have discussed rest squarely on the *same-property* interpretation. Consequently, any challenges faced by the latter will affect the former.

4. Criteria for Interpretations of Einstein's Equation

The familiar goal of philosophical interpretations of physical theories is to answer the question 'What would the world be like if this theory were true?'. Typically, one assumes, often implicitly, that the answer to such a question satisfies the following additional criteria. First, an interpretation *I* of a given physical theory *T* does not appeal to hypotheses outside *T* or theories other than *T*. For example, interpretations of elementary quantum mechanics do not assume hypotheses from any other theory. Of course, this does not preclude subsequent explorations concerning how *I* is affected by other theories. Second, *I* is either *philosophically uniform* or provides compelling grounds for any non-uniformity. To say that *I* is *philosophically uniform* is just to say that either *I* treats elements in the mathematical formalism of *T* that are similar in type (very roughly speaking) on a par, or *I* explains why *some* of these elements ought to be treated differently. This notion is difficult to make precise. However, we have some clear cases to guide our intuitions. For example, if an interpretation of a spacetime theory holds that coordinate systems are artifacts used for our purposes and are *not* genuine parts of 'reality', then we expect the interpretation to treat *all* coordinate systems in this way unless compelling reasons are given for treating some coordinate systems differently. Similarly, if an interpretation of special relativity holds that Lorentz-invariant quantities represent objective features of physical systems, then we expect this to apply to all Lorentz-invariant quantities, unless compelling reasons are given for treating some Lorentz-invariant quantities differently.

One can adapt these general requirements on interpretations to our specific purpose. A *viable* interpretation of Einstein's equation *I* should satisfy the following criteria:

(I1) *I* answers the question 'What would the world be like if special relativity, and specifically Einstein's equation, were true?'
(I2) *I* does not appeal to the truth of any hypotheses from theories *other than* special relativity.
(I3) The interpretation of special relativity upon which *I* is based is *philosophically uniform* or compelling reasons are cited for adopting any non-uniformity in the interpretation of special relativity.

I take it that (I1)–(I3) are largely uncontroversial. Nevertheless, I wish to note the following. First, (I2) is a methodological constraint that simply confines our task to interpreting one physical theory at a time. The hope is that we achieve greater clarity by proceeding along these lines and only subsequently examining the relationships among different theories. However, (I2) might be construed as unnecessarily restrictive. One might argue that our philosophical goal ought to be to understand what the world is like if modern physics as a whole is true. This would not be a serious objection. One can always attempt to attain a broader understanding of the philosophical consequences of modern physics by proceeding first along the lines suggested by (I2). Furthermore, since (I2) is merely a methodological requirement, interpretations of Einstein's equations that fail to satisfy (I2) need not be rejected outright. Such interpretations may be considered so long as the additional hypotheses being assumed are clearly articulated, the additional hypotheses are consistent with special relativity, and there is compelling evidence for truth of such hypotheses. Second, interpretations that fail to satisfy (I3) face a far more serious challenge than those that merely fail (I2). An interpretation that fails (I3) is undermined because it is based on an interpretation of special relativity that contains an unjustified degree of arbitrariness. Third, (I1)–(I3) are only intended as requirements for a *viable* interpretation; they are not intended to select a preferred interpretation of Einstein's equation. Thus, in the next section (Section 4), I proceed by first determining which of the interpretations we have canvassed satisfy (I1)–(I3).

5. Viability of Interpretations of Einstein's Equation

Of the four property interpretations of mass–energy equivalence, the one that faces the most serious challenges, since it fails to satisfy (I3), is Lange's *one-property, no-conversion* interpretation. As we have seen, Lange argues that the rest-mass of a body is a real, objective property, whereas the energy of a body is not. To establish this claim, Lange makes a fairly simple argument. Lange first invokes, in several places, invariance under the relevant group for a given spacetime theory as a *necessary* condition for a quantity to represent a 'real' feature of nature. For example, Lange says, 'a real quantity must be invariant' (Lange 2002, 206). In the case of special relativity, Lange uses a simple *modus tollens* to exclude any non-Lorentz-invariant quantity from the set of quantities that designate 'real' properties. Significantly, Lange argues energy is not real because energy, by which he typically means kinetic energy *simpliciter*, is not Lorentz-invariant. For Lange, non-Lorentz-invariant quantities fail to be 'real' because they fail to represent 'the objective facts, on which all inertial frames agree' (Lange 2002, 209). Non-Lorentz-invariant quantities are tainted with the particular 'perspective' of a given inertial frame.

However, Lange uses Lorentz-invariance not only as a necessary condition, but also as a *sufficient* condition for a quantity to be real. For example, in his discussion of length, Lange states:

> Though a body's length differs in different frames, and so is not Lorentz-invariant, a body's length *in a given frame* is the same in all frames. (This quantity carries its reference to a particular frame along with it, so to speak.) Therefore, this quantity is objectively real. (Lange 2002, 218)

The inference to the conclusion that a body's length in a given reference frame is real is valid only if Lange assumes that Lorentz-invariance is a *sufficient* condition for the 'objective reality' of a quantity. Lange makes this assumption in several places, for example, when he argues for the reality of the Minkowski interval (Lange 2002, 219), the time order of time-like separated events (Lange 2002, 219–20), a system's total mass (Lange 2002, 223), and mass more generally (i.e. the mass of either a body or a system of bodies) (Lange 2002, 225). All of these features of physical systems or events in spacetime are real *because* they are Lorentz-invariant. For example, Lange says, 'mass is a real property (since it is Lorentz-invariant)' (Lange 2002, 227). Energy is not a real property *because* it is not Lorentz-invariant. Thus, Lange reaches the core of his *one-property* interpretation by using Lorentz-invariance as both a *necessary* and *sufficient* condition for a property to be real.

Although there is no *prima facie* problem with relating Lorentz-invariance to the 'reality' of quantities, the challenge for Lange is that he does not apply Lorentz-invariance as a *sufficient* condition uniformly, and he offers no grounds for the non-uniformity he introduces. If, as Lange claims, he is focusing on the equivalence of rest-mass and rest-energy (Lange 2002, 224–25), then he is focusing on the equivalence of two scalar invariants. If invariance is a sufficient condition for a quantity to be real, then rest-energy must be a real quantity in precisely the same sense that rest-mass is a real quantity. Furthermore, Lange does not explain why he treats rest-mass differently from rest-energy. Thus, Lange's interpretation seems to fail to satisfy (I3).

If my observations are correct, Lange seems to have the following two options for repairing his non-uniform interpretation of Lorentz-invariant quantities. First, Lange might reply that invariance is just one of a collection of conditions that jointly suffice to show that a quantity is real. However, it is not very clear what those other conditions *could* be. According to Lange, rest-mass is real, because it is invariant. Yet, for Lange, other scalar invariants, specifically proper length and rest-energy, are *not* real. Furthermore, his reasons for rejecting proper length as real are different from his reasons for rejecting energy as real. Proper length is not real, according to Lange, because although it is invariant, it is merely one of indefinitely many lengths all of which are Lorentz-invariant. We have, Lange claims, no grounds for selecting proper length as *the* real length of an object (p. 218). Lange does not explain why this argument does not apply, *mutatis mutandis*, to mass. When it comes to energy, Lange seems to slip from discussing total energy and kinetic energy *simpliciter*, which are not invariant, to discussing rest-energy. Thus, it is difficult to see, prima facie at least, what additional condition or set of conditions Lange could provide to single out rest-mass as one of the invariants that counts as real, along with the Minkowski interval and the time order of time-like separated events. After all, Lange would want the same condition or set of conditions to rule out as real other invariants such as proper length and rest-energy.

The second option Lange seems to have is to adhere to the view that invariance is a *necessary* but not *sufficient* condition for a quantity to be real.[1] However, this is at odds with the familiar observation he adopts that *if* a quantity is invariant, *then* all observers agree on its value, and hence it represents an objective and real feature of the world. Furthermore, if Lange did not use invariance as a *sufficient* condition, then he would have to provide us with a different condition, or set of conditions, for supposing that rest-mass designates a real quantity. Again, it is not clear what that condition, or set of conditions, could be, especially since the conclusion Lange is aiming for is that rest-mass is real, whereas rest-energy and proper length are not. Because neither of the options I have considered seems promising, it seems Lange's interpretation of Einstein's equation is not viable as it fails to satisfy (I3). Furthermore, if we try to impose a uniform interpretation of invariance on Lange's interpretation, then we are led to the conclusion that either rest-mass and rest-energy are both real, or neither one is. In the former case, the question concerning whether mass and energy are the same property and whether they are 'interconvertible' is entirely open.

Even if my observations about Lange's interpretation are correct, they only seem to affect his claim that mass is a real property, but energy is not. Lange's arguments against the view that there is a sense in which mass can be converted into energy (or vice versa) seem independent of his *one-property* view. For example, Lange argues that when a gas sample is heated, its mass increases according to Einstein's equation. We are thus tempted to say that energy has been 'converted' into mass. However, if we analyze the gas at a microscopic level, we see that the only physical process occurring is that the heat energy of the source is being transformed into the kinetic energy of the molecules of the gas. From a microscopic perspective, there is no 'conversion' of mass into energy. Thus, Lange concludes that *we* 'convert' energy into mass when we shift from analyzing a system from the microscopic to the macroscopic level.

That energy is not a real property simply does not seem to enter the argument. It seems Lange's argument could easily be made by someone who believes energy, specifically rest-energy, is a real property. Such an argument would have to include a definition of the rest-mass of the gas sample as a function of the dynamical variables of the constituent molecules. However, this is easily done. Nevertheless, regardless of whether we regard rest-energy as real, Lange's conclusion seems a bit hasty.

Lange agrees that when the gas sample is heated, its inertial mass increases. Where does this additional inertial mass come from? It is as if Lange tried to 'open up the black box' and look inside the gas sample to find the additional mass. When we do this, we do not find that the gas sample contains more *matter*. We only find that the constituents of the gas sample have an increased amount of energy. Thus, we conclude that the increased energy at the microscopic level somehow manifests itself as inertial mass one level up. However, it does not follow from this that *we* somehow 'converted' energy into mass by shifting perspectives. Even if no human observer examined the sample of gas or even theorized about it, in any interaction where it responds as a single body (e.g. if the container enclosing the gas sample is struck by another object), the gas sample responds with a greater inertial mass after it is heated. Furthermore, if, contra Lange, rest-energy is a real property, then Lange's example is just the kind of example Bondi and Spurgin use to support their *no-conversion* interpretation: there is no genuine 'conversion' of energy into mass. All we have is a transformation of one kind of energy, the heat energy of the source, into another kind of energy, the kinetic energy of the molecules. This additional energy of the constituents of the gas sample *contributes* to the rest-energy of the gas and hence, through Einstein's equation, to its rest-mass. Precisely why or how the energy of the constituents of a body contributes to that body's rest-mass is a question to which I shall return in the next section.

Whereas the main challenge to Lange's *one-property, no-conversion* interpretation comes from (I3), the main challenge to the *same-property, no-conversion* interpretation comes from (I2). As a number of authors have observed, e.g. Rindler (1977) and Stachel and Torretti (1982), nothing in special relativity rules out the possibility that there exists matter that cannot radiate all of its mass in the form of energy. It is consistent with Einstein's equation that there exists a certain kind of matter, call it 'exotic matter', all of whose mass is inert, in the sense that it can *never* be radiated away as energy. To see this, one merely needs to observe that the relation that one actually derives from the two postulates of special relativity is that $E = (m - q) + K$ (in units in which $c = 1$).[2] K is an additive factor that is routinely set to zero and merely fixes the zero point of energy. The term q is also routinely set to zero. However, setting q to zero involves a hypothesis concerning the nature of matter. Specifically, it amounts to adopting the hypothesis that exotic matter does *not* exist. Yet, the view that mass and energy are the *same* property seems to require that we set q to zero, for if there were matter that had mass that was not 'convertible' into energy, then it would seem that mass and energy could not be the *same* property after all. Thus, the *same-property, no-conversion* interpretation of Einstein's equation violates (I2), because it *requires* that we adopt a hypothesis concerning the nature of matter that lies *outside* special relativity.

As I observed earlier, the violation of (I2) is not nearly as serious as the violation of (I3), which the *same-property* interpretation does not violate. In this particular case, all we need for the *same-property*, *no-conversion* interpretation to be viable is compelling justification for setting q to zero, as such a hypothesis is clearly consistent with special relativity. The available evidence is of two sorts. On the one hand, we have not yet found any matter for which the value of q is non-zero; we have found no exotic matter. On the other hand, we have found convincing cases of matter for which q is equal to zero, for example, in annihilation collisions where the entire mass of the two incoming particles becomes energy. The evidence is compelling though, of course, not conclusive. Consequently, if we adhere to our criteria for interpretations of Einstein's equations (I1)–(I3) closely, we must conclude that the *same-property* interpretation is not a viable interpretation of mass–energy equivalence. If, on the other hand, we are willing to allow one additional hypothesis, which is fairly well confirmed, then the *same-property*, *no-conversion* interpretation is viable. Finally, since both of the ontological interpretations I have considered rest on the *same-property* interpretation, the viability of the former is tied to the viability of the latter.

Surprisingly, perhaps, both of the *different-properties* interpretations are also closely related to hypotheses concerning the nature of matter, though in significantly different ways. The *different-properties, no-conversion* interpretation requires that we make a commitment to a hypothesis concerning the analyzability of matter. According to the *different-properties*, *no-conversion* interpretation, purported cases of conversion of mass and energy are cases where there is merely a transfer of energy. For example, consider an interaction among atomic and subatomic particles of the sort ordinarily used in textbooks to illustrate the 'conversion' of mass and energy. According to the *different-properties, no-conversion* interpretation, none of the mass of the reactants is strictly speaking 'converted' into the kinetic energy of the products (say). Instead, some of the energy of the constituents of the reactants is transformed into the kinetic energy of the products. Thus, this interpretation assumes that particles engaged in such collisions are always *composite* particles. If, as Bondi and Spurgin suggest, one were to adopt the *different-properties*, *no-conversion* interpretation as *the* interpretation of Einstein's equation, then it seems one would have to adopt the hypothesis that matter is *always* analyzable into constituent parts. Thus, the *different-properties, no-conversion* interpretation also violates (I2), as a hypothesis concerning the analyzability of matter is clearly outside the scope of special relativity.

As in the case of the *same-property* interpretation, one might want to weaken (I2) to allow one hypothesis external to special relativity. Only this time, it is not so clear that we have compelling evidence that warrants our additional hypothesis, for what we seem to need is evidence for the infinite analyzability of matter. Furthermore, the existence of particles, such as the electron, which seem to be unanalysable and which participate in annihilation reactions, speaks against the infinite analyzability of matter. Thus, the *different-properties*, *no-conversion* interpretation of mass–energy equivalence fails to be a viable interpretation.

The *different-properties*, *conversion* interpretation of Einstein's equation is the only one that satisfies (I1)–(I3) without requiring that one modify the criteria. (I1) is satisfied trivially. (I2) is satisfied because the *different-properties*, *conversion* interpretation stays within the confines of what one can derive *within* special relativity without invoking additional hypotheses. This interpretation does not state that mass–energy 'conversions' must occur. It only states that *if* the entire mass of, say, two particles 'disappears' when they collide, then an equivalent amount of energy must 'appear' somewhere in the system. Notice that this requires neither that we set $q = 0$ nor that we adopt any additional hypotheses concerning the nature of matter. Finally, the *different-properties*, *conversion* interpretation satisfies (I3) because it adopts a uniform interpretation of the mathematical formalism of special relativity. Specifically, in contrast to Lange's interpretation, the *different-properties*, *conversion* interpretation treats all scalar invariants as representatives of real features of physical systems. Thus, the *different-properties*, *conversion* interpretation is a viable interpretation of Einstein's equation.

6. Preferred Interpretation of Einstein's Equation, Types of Explanations, and Types of Physical Theories

Our choice concerning which interpretation of Einstein's equation to prefer is rather constrained given the results of the last section. If we adhere strictly to (I1)–(I3), then it seems we must prefer the *different-properties*, *conversion* interpretation, since it is the only viable interpretation of Einstein's equation we have discussed. We have not, of course, ruled out the possibility that one could frame a different viable interpretation of $E = mc^2$, though there do not seem to be too many options left. Opting for the *different-properties*, *conversion* interpretation has its merits, as I will soon discuss. However, I first want to consider weakening (I2) to make the *same-property* interpretation viable. I will argue that the *same-property*, *no-conversion* interpretation is not the interpretation of $E = mc^2$ we ought to prefer.

Let us grant that we have relaxed (I2) and that we have compelling evidence for setting $q = 0$ in $E = m - q$, i.e. we have compelling evidence that there is no exotic matter. On this hypothesis, mass and energy are always numerically equal and can be expressed in the same units. According to the *same-property* interpretation, it follows that they are the same property. However, as Bondi and Spurgin implicitly suggest, the conclusion seems a bit hasty. For Bondi and Spurgin, mass and energy are different properties because they have different *dimensions*. Their claim is not about the *units* one uses for mass and energy. Thus, Bondi and Spurgin are suggesting that two properties that can be measured in the same units need not have the same dimensions. Reflecting on how one selects units in which $c = 1$ bears out Bondi and Spurgin's suggestion.

Selecting units in which $c = 1$ amounts to performing a substitution of variables. Instead of using coordinates (x, t) we use coordinates (x^*, t) where $x^* = x / c$.[3] The variable x^* has units of time, and it has dimensions of time. When we specify a certain value of x^*, we have specified an amount of time. Nevertheless, since x^* indicates a *distance*, and one cannot specify a length using a time without specifying a velocity, we preface the time units of x^* with the expression 'light', as in 'light-years'. We are not merely appeasing our jarred intuitions. Instead, by using 'light-years' (say), we want to indicate that to recover a *distance*, i.e. a quantity with dimensions of *length*, we have to multiply x^* times c. Consequently, although x^* and t are expressed in the same units, and have the same dimensions, it does not follow that x and t have the same dimensions. To say that 'we can measure distance in units of time' is simply to say that there are contingent facts that make it possible to perform the substitution of variables defined by $x^* = x / c$.

When we use coordinates (x, t) and standard units, e.g. meters for x and seconds for t, we use three fundamental dimensions in dynamics, viz., length (L), time (T), and mass (M). Energy has derived dimensions of $M(L^2 / T^2)$. However, if we use units in which $c = 1$, we use only two fundamental dimensions, M and T. In this case, what is called 'energy', which I will designate with E^*, has dimensions of mass. However, just as in the case of space and time, it does not follow that E and m have the same dimension. A mass is simply not an energy until we multiply it times a velocity squared. To emphasize this, one might have chosen to express energies E^* in units prefaced by the awkward expression 'light-squared', such as 'light-squared-kilograms', to retain the parallel with space and time. Again, this would be not merely to appease our jarred intuitions but to underscore that energy does not have the same dimensions as mass and that to recover an energy from a mass, we have to multiply times c^2. In effect, in selecting units in which $c = 1$, we have also made the substitution of variables defined by $E^* = E / c^2$.

Yet another way to make Bondi and Spurgin's point is to say that energy would have the same dimensions as mass if, and only if, lengths and durations had the same dimensions, i.e. if L and T were the same dimension, or at least dimensions of the same type. However, there are well-known reasons for not regarding time as a spatial dimension. Nothing in special relativity forces us to erase the distinction between space and time in this sense, despite our ability to express lengths in temporal units by making the substitution of variables I have described above. Furthermore, none of this is affected by whether one adopts the four-dimensional co-variant formulation of special relativity or the coordinate-transformation approach. To make the point in yet a different way, the substitution of variables that underwrites our choice of units in which $c = 1$ does not, strictly speaking, entail that c is a dimensionless quantity in the same sense that π is a dimensionless quantity. Thus, even if we relax (I2), I think there are good reasons for not adopting the *same-property* interpretation of mass–energy equivalence. Finally, this has the consequence that we should hesitate before we adopt either *ontological* interpretation of Einstein's equation.

Our only choice, then, seems to be to prefer the *different-properties*, *conversion* interpretation. I now want to argue that this is not a bad choice. Clearly, the *different-properties*, *conversion* interpretation does not answer the questions that some have asked of Einstein's equation. For example, Lange seems concerned with answering the question 'Why does energy at one level in the analysis of matter manifest itself as mass one level up?' He then searches for an answer by attempting to frame a bottom-up explanation, i.e. an explanation of the macroscopic properties of physical systems in terms of the properties of their constituents. The *different-properties*, *conversion* interpretation simply remains silent about Lange's question. However, this is a merit of the *different-properties*, *conversion* interpretation, for it indicates that this interpretation recognizes that Lange's question is ill posed. Special relativity cannot provide bottom-up explanations for any of its consequences.

As Einstein (1919, 228) pointed out, special relativity is what he called a *principle* theory. It is a theory that begins with two principles that entail a set of constraints, such as $E = mc^2$, that all physical systems must satisfy. Significantly, special relativity is not a *constructive* theory, i.e. a theory that describes the behavior of matter by appealing to its constituents. One of the merits of the *different-properties*, *conversion* interpretation is that it does not blur this distinction. So, for example, in contrast to Lange, a proponent of the *different-properties*, *conversion* interpretation would not think to try to 'open up the black box' to discover how mass is converted into energy. As I have argued elsewhere (Flores 1999), principle theories only offer top-down explanations. Consequently, while special relativity might tell us *that* mass and energy can be converted in certain circumstances in the sense that a certain amount of mass (say) 'disappears' and an equivalent amount of energy 'appears', it does not tell us *why*, if the answer we are looking for is an explanation of the behavior of a composite system in terms of its parts. Similarly, while relativity tells us *that* the energy of the constituents of a system of particles *contributes* to the mass of the system, it does not tell us *why*, again if by asking 'why?' we mean to find an explanation of an object in terms of its constituents. Principle theories do not afford bottom-up explanations (Flores 1999).

Finally, a constructive theory cannot explain why energy at one level manifests itself as mass one level up precisely because if such a theory is relativistic, it must already *assume* mass–energy equivalence. As long as Einstein's equation is assumed as a principle, or derived from a set of principles, no constructive theory constrained by such principles can offer a bottom-up explanation for mass–energy equivalence. So, for example, not even string theory, if successful, will explain why $E = mc^2$, if we are hoping for a bottom-up explanation. Instead, string theory seems to take Einstein's equation as axiomatic and then to use it to explain how different vibrational modes of strings can account for the observed masses of subatomic particles. And if we ask why the vibrational energy of a string manifests itself as mass one level up, the only answer we find is that it is because of Einstein's equation. Beyond that, string theory is, and must remain, entirely silent.

7. Conclusion

Most interpretations of Einstein's equation in the literature fail the methodological requirement that interpretations of physical theories ought to restrict their claims to the theory being interpreted. For example, both the *same-property*, *no-conversion* and the *different-properties*, *no-conversion* interpretations make assumptions concerning the nature of matter. The only extant interpretation that does not make such assumptions, and which does not face other challenges in the way that Lange's *one-property*, *no-conversion* does, is the *different-properties*, *conversion* interpretation. According to the latter, mass and energy are distinct properties, many purported cases of 'conversion' are merely transfers of energy, but there may be cases of 'conversion' where an amount of mass (say) 'disappears' and a corresponding amount of energy 'appears'. Thus, strictly speaking, if our methodological requirement is well motivated, then we ought to prefer the *different-properties*, *conversion* interpretation of Einstein's equation. Finally, even if we chose to relax our methodological requirement, which makes the *same-property*, *no-conversion* interpretation viable, I have argued that we ought to prefer the *different-properties*, *conversion* interpretation. The *same-property*, *no-conversion* interpretation fails because it makes an illegitimate inference from our ability to express two quantities in the same units to the conclusion that they are therefore the same. Furthermore, I have argued that allowing hypotheses concerning the nature of matter to influence our interpretation of mass–energy equivalence blurs the distinction between Einstein's principle and constructive theories, which if kept separate allows us to understand why mass–energy equivalence only receives a top-down explanation.

Acknowledgements

I would like to thank James W. McAllister, the editor, and two anonymous referees for their careful and constructive critique of an earlier version of this essay. For her insightful and constructive criticisms on all versions of this essay, I would like to thank Rachel Fern.

Notes

[1] In a recent communication to the author, Lange claims that in his work, he only uses Lorentz-invariance as a necessary condition. I believe evidence from his text indicates otherwise. However, Lange may *now* wish to restrict himself to treating Lorentz-invariance as a necessary condition.

[2] Although seldom noted in textbooks, the full relation $E = (m - q) + K$ is explicitly derived in many 'purely dynamical' derivations. See, for example, Mermin and Feigenbaum (1990).

[3] For simplicity, I am assuming we are working with only two dimensions.

References

Bondi H., and C. B. Spurgin. 1987. Energy has mass. *Physics Bulletin* 38: 62–63.

Eddington, A. S. 1929. *Space, time, and gravitation.* Cambridge: Cambridge University Press [originally published in 1920].

Einstein, A. 1905. Does the inertia of a body depend upon its energy content? In *The principle of relativity,* edited by A. Einstein, H. Minkowski & H. Weyl (1952). New York: Dover, pp. 69–71.

Einstein, A. 1919. What is the theory of relativity? In *Ideas and opinions,* edited by A. Einstein (1982). New York: Crown, pp. 227–232.

Einstein, A., and L. Infeld. 1938. *The evolution of physics.* New York: Simon & Schuster.

Flores, F. 1999. Einstein's theory of theories and types of theoretical explanation. *International Studies in the Philosophy of Science* 13: 123–33.

Lange, M. 2001. The most famous equation. *Journal of Philosophy* 98: 219–38.

Lange, M. 2002. *An introduction to the philosophy of physics.* Oxford: Blackwell.

Mermin D., and J. Feigenbaum. 1990. $E = mc^2$ in *Boojums all the way through.* New York: Cambridge University Press.

Rindler, W. 1977. *Essential Relativity.* New York: Springer Verlag.

Stachel, J., and R. Torretti. 1982. Einstein's first derivation of mass–energy equivalence. *American Journal of Physics* 50: 760–61.

Torretti, R. 1996. *Relativity and geometry.* New York: Dover.

Zahar, E. 1989. *Einstein's revolution: a study in heuristic.* La Salle, IL: Open Court.