Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 31, No. 2, pp. 187–199, 2000

1355-2198/00 $ - see front matter

Energy Conservation in GTR

*Carl Hoefer**

The topics of gravitational field energy and energy-momentum conservation in General Relativity theory have been unjustly neglected by philosophers. If the gravitational field in space free of ordinary matter, as represented by the metric g_{ab} itself, can be said to carry genuine energy and momentum, this is a powerful argument for adopting the substantivalist view of spacetime.

This paper explores the standard textbook account of gravitational field energy and argues that (a) so-called stress-energy of the gravitational field is well-defined neither locally nor globally; and (b) there is no general principle of energy-momentum conservation to be found in General Relativity. I discuss the nature and justification of the zero-divergence law for ordinary stress-energy, and its possible connection with the failure of General Relativity to realise Mach's principle.

Keywords: Einstein; Energy Conservation; Spacetime; Gravity Waves; Substantivalism; Mach's Principle

1. Introduction

The topic of this paper is energy-momentum conservation in classical GTR (the General Theory of Relativity). This is an area that has received almost no attention at all from philosophers. What I hope to show is that this neglect is very much unjustified, and that there are important conceptual and interpretive issues deeply connected with energy-momentum conservation (or the lack of same) in GTR. It is a crucial interpretive issue bearing on the ontology of

* Department of Philosophy, Logic and Scientific Method, The London School of Economics, Houghton Street, London WC2A 2AE, U.K. (*e-mail*: c.hoefer@lse.ac.uk).

PII: S1355-2198(00)00004-6

spacetime and the relationist/substantivalist debate. It is also likely to be deeply linked to the question of how to reconcile GTR with quantum mechanics.[1]

The issue of conservation laws is connected to the debate between substantial and relational views of spacetime in at least two ways. First, the energy-momentum tensor T^{ab} represents the 'matter fields', the only 'stuff' that *really* exists according to the relationist. But if $T^{ab} = 0$ 'empty space' can carry *genuine* energy-momentum of the gravitational field, then it (the empty space) should be counted as real also, and spacetime itself as represented by g^{ab} should be considered substantial and real. John Earman likes to stress the reality of such gravitational stress-energy in gravity waves, remarking that since it can knock down walls and light light-bulbs (in principle), it is certainly real enough. Or is it?

The key here is of course the word 'genuine', and what we will see is that the quantities representing energy-momentum of the gravity field probably do not deserve the label. A second, more speculative link with the ontological debate is the following: the imposition of 'the law of conservation' in GTR may be connected with the failure of that theory to satisfy Mach's Principle.

Let me begin by stating one of my main claims, which pertains to what is widely described as *the* law of energy-momentum conservation in GTR:

$$\nabla \cdot \boldsymbol{T} = 0, \qquad T^{ab}{}_{;b} = 0, \tag{1}$$

where $\nabla \cdot T$ is the covariant divergence of $\mathbf{T}$, the stress-energy tensor of matter, in intrinsic notation, and $T^{ab}{}_{;b}$ is the same quantity in coordinate-based notation.[2] For all its elegance and beauty, I claim, this is *no* genuine conservation law.

In most of the rest of the paper I will begin to explore the issues surrounding this claim. A full treatment of either the conceptual or the historical questions being raised will be beyond the scope of this paper, but I hope to say enough about the issues to make it clear that this topic ought to be of central importance to philosophers and historians of physics.

2. Some Background

Conservation of mass and conservation of energy are often thought of as deeply enshrined principles of classical (pre-GTR and pre-quantum mechanics) physics. In fact, however, the tenure of the latter as a well-established idea was arguably fairly short (limited to part of the nineteenth and early twentieth centuries), and at all times fraught with difficulties. The particular difficulty of most interest to us is that of energy conservation in Newtonian gravity theory.

[1] See Callender and Huggett (2000) for a set of recent discussions of philosophical problems in quantum gravity.

[2] Semicolons indicate covariant derivatives, commas indicate ordinary derivatives, and the Einstein convention (indices repeated are to be read as summed over) is followed throughout the paper.

To put it plainly, the energy of a system is patently *not* conserved when gravity is at work. Consider a system of two masses at relative rest, separated by distance d. There is zero net momentum in their CM (centre of momentum) frame, and their total energy in classical terms comprises their kinetic energy of motion (zero) and their heat content.[3] Supposing the masses are large and the distance fairly great, when gravity accelerates the two masses together a violent collision will ensue. Afterward, then, we will still have zero net momentum and kinetic energy of motion in the CM frame; but the heat content will have gone up quite a bit!

This is an unfair way to put it, of course; everyone knows that the extra energy came from gravity, and we should say that we were forgetting the *potential energy*:

$$U(r) = Gm_1 m_2/r, \tag{2}$$

the total energy $E_k + U$ being what is conserved.

But notice: the gravitational energy is not localised, and unlike other forms of energy in classical physics, one has no physical story about a *local interaction* that transforms this alleged form of energy into heat and motion. So gravitational potential energy is often described in freshman physics texts as a mere 'bookkeeping device'. But at least it does have the virtue of genuinely *balancing* the books—which is precisely what gets lost in GTR, as we will see later.

In classical field physics (fluids, electromagnetism) we have also the local differential conservation laws:

$$T^{ab}{}_{,b} = G^a, \tag{3*}$$

where G^a is the external force density (gravitational).

In the *absence* of gravity, we have

$$T^{ab}{}_{,b} = 0. \tag{3}$$

Notice the crucial comma here instead of semicolon; it is the ordinary divergence, not the covariant divergence, whose vanishing represents a conservation principle. What these equations tell us is that locally, no rest mass or momentum is being created, unless it is being created by a gravitational force. Gauss' theorem then lets us turn (3) into an integral conservation law:

$$\int_V T^{ab}{}_{,b} d^4V - \int_S T^{ab}\, d^3 \sum_b = 0, \tag{4}$$

where in the first term we integrate over a 4-volume, and in the second over a 3-surface bounding that volume (with implicit surface-normal one-form n_b). The

3 I here exclude binding energies (including of course gravitational binding energy, which we will come to shortly).

upshot of such an integral conservation law is that energy and momentum are not changing within the volume except by crossing the surface bounding it.

Again, it should be borne in mind that this equation applies only in the absence of gravity. There is no inclusion of gravitational potential energy in T^{ab}, and the simple 2-mass system we considered above would violate this law. (Though, again, the books can be balanced by appropriately including the gravitational force.)

There is precedent, then, for thinking of gravity as something that either violates energy conservation principles, or satisfies them only by virtue of a not-wholly-satisfying mathematical artifice. Despite this, insistence on conservation of energy-momentum played a historically crucial role in the genesis of GTR.

The history of the roles played by conservation principles in Einstein's early work is a tangled story to unravel. It seems that the certainty that one or more conservation laws should be central to the new gravity theory played two distinct roles for Einstein, often simultaneously. First, it could be used as a heuristic guide in trying to find the correct field equations for gravity, from 1913 to 1915. Second, satisfaction of a nice-looking conservation law was a point that could be cited in support of field equations, even when they were arrived at by another route.[4]

Some historical questions will re-surface in the last section, where I evaluate Einstein's approach to energy-momentum conservation in light of his Machian goals for GTR. For now I will turn directly to what I will call the 'received view' on energy conservation in GTR. As it happens, this view was quite literally received, *by* the physics community *from* Einstein himself, essentially complete in 1916.[5]

3. The Received View

Equation (1) is frequently justified in textbooks along these lines (e.g. Schutz, 1985, p. 185):

> (1) is the proper energy conservation law for GTR because (3) is the correct classical law, and the 'comma goes to semicolon' rule tells us that (1) is therefore the correct equation in the curved spaces of GTR.

The 'comma goes to semicolon' rule essentially tells you that if an equation valid in classical physics (which is of course set in a flat spacetime) involves partial

[4] The latter role finds a clear example in the 1916 exposition of the full theory, in which Einstein uses satisfaction of classical-looking conservation equations as evidence for the theory being correct, first about the matter-free field, then later about the general field equations. See Einstein (1916, Section 16).

[5] Einstein (1916). There were early challenges to gravitational waves as real energy-bearers, and to Einstein's view of conservation, made by Schrödinger and Levi-Civita, but the physics community as a whole quickly adopted Einstein's views. See Cattani and De Maria (1989).

derivatives, divergences etc., these should be replaced by *covariant* derivatives, divergences, etc. to get the correct GTR equivalent. At the origin of a local inertial (freely falling) reference frame, the semicolon-bearing equations will reduce to the originals. The comma-to-semicolon rule is therefore usually presented as a consequence (or even a re-expression) of the equivalence principle (see e.g. Misner *et al.*, 1973, p. 387). By 'equivalence principle' I mean what is usually called the 'Strong Equivalence Principle':

> **Strong Equivalence Principle (SEP)**: At the origin of a local inertial reference frame (LIF), all physical laws are the same as their classical counterparts in special relativity.

This statement is rough, but typical; when one tries to make it more precise, the SEP tends to blossom into several distinct parts or sub-principles.[6]

We will return shortly to the question of what justifies Equation (1). At first it looks like the most natural move to make, justified by its elegance and economy as well as the conviction that there should be *some* energy-momentum conservation law. But the problem is that, contradicting this line we have just rehearsed (and very often in the same text, a tradition started by Einstein), we find that another part of the received view is this (for example Einstein, 1916, p. 151):

> (1) is not *really* a conservation law (which must involve ordinary, not covariant divergence); properly speaking there is no local law of energy-momentum conservation in GTR. Instead, (1′) *specifies the interchange of energy and momentum between gravitational and ordinary stress-energy*, where we write out (1) explicitly:

$$T^{ab}_{;b} = T^{ab}_{,b} + \Gamma^{a}_{cb}T^{cb} + \Gamma^{b}_{cb}T^{ac} = 0. \qquad (1')$$

The apparent contradiction here is striking. On the one hand, (1) is a perfect and proper conservation law for energy-momentum. Comma has gone to semicolon, that is all. On the other hand, actually, it is not a conservation law after all, but rather a replacement for (3*), something that describes how energy coming from outside (gravitational energy) changes the balance locally. Equation (1) is conceded not to be a proper conservation principle because it cannot be used to write an integral conservation law such as (4).[7] Intuitively, if energy-momentum is *really* being conserved locally, then when one integrates up it should be conserved over regions as well. Since this fails, we have to fall back on a different understanding of what (1) represents.

[6] See for example the discussions of equivalence principles in Weinberg (1972) or Rindler (1977).

[7] Integration is the opposite of ordinary (partial) differentiation, not covariant differentiation, so Gauss' theorem cannot be applied to (1).

What then of our comma-to-semicolon rule and its justification by SEP? The correct thing to say, it seems to me, is that according to the received view this justification was tenuous, but happened to give the right results anyway.

First, let us refine the SEP as Weinberg and others do, by distinguishing a Medium-Strong EP and a Very-Strong EP (Weinberg, 1972, p. 70). The Medium-SEP says that the laws of all *non-gravitational* physics have their STR forms at the origin of a LIF. This part of the SEP enjoys great intuitive plausibility as well as fairly strong experimental support. The intuitive justification is this. Laws of non-gravitational physics in classical physics tend to be stated in forms that are only valid in inertial frames. When one moves to a covariant formulation of such laws, one has to change commas to semicolons just to ensure that the laws say the same thing. These laws, presumably, are known to be valid (to a good approximation) in the *flat* spacetimes of classical physics. The physics only involves 'coupling' to the spacetime structure in the sensitivity to affine structure displayed in the laws' covariant formulations.

Now, the leap we make with the Medium-SEP is the assumption that, when spacetime is *curved*, the physics still only involves coupling to the affine structure, in the same way as before. What other possibility is there? At least one, namely *curvature coupling*: that the correct laws should contain terms explicitly involving the Riemann curvature tensor (or one of its contractions). Curvature coupling could go unnoticed in regions where spacetime is flat or nearly so, and yet be important when curvature is significant. The Medium-SEP rules out curvature coupling for non-gravitational physics.

But (1) is clearly *not* non-gravitational physics, and it requires the Very-SEP to justify the application of comma-to-semicolon that produces it.[8] And the justification for the Very-SEP is weaker, both intuitively and experimentally. Intuitively the point should be obvious. If the Very-SEP is saying that there is no curvature coupling in *gravitational* physics, it is unclear what this could mean while still having a chance to be true. Einstein's equations, after all, are in one sense precisely the specification of *how* gravitation is coupled with (or, produces) curvature. And generally speaking, the flow of stress-energy—which is the topic of a differential conservation law—is taken in GTR both to *affect* spacetime curvature (the field equations), and to *be affected* by it (the law of motion). Experimentally, as some texts point out, it would be nice to have direct tests of the Very-SEP, but it is not clear that we do.[9]

[8] As a presumed replacement of the classical energy conservation principle in the *absence* of gravity, one might suppose that (1) should count as non-gravitational physics. But this is a mistake, as the talk of (1) being about the interchange of energy and momentum between gravitational field and matter shows. It is also obvious that, in a context of trying to decide whether T^{ab} is locally conserved when gravity is non-negligible, it begs the question to count this as a matter of non-gravitational physics!

[9] Weinberg (1972) describes one possible such test, currently not realisable, and remarks: 'This is a pity, because it is precisely the very strong assumption that the Principle of Equivalence applies to gravitational fields that will lead us in Chapter 5 to Einstein's field equations for gravitation' (p. 70). Weinberg is presumably referring to the key role that assuming the truth of equation (1) plays in his derivation of Einstein's field equations. See also Schutz (1985, p. 184).

So to apply the comma-to-semicolon rule to the conservation law is in effect to make a particular, strong assumption about the precise nature of the relationship between stress-energy flow and curvature (call this the assumption of 'minimal coupling'. The assumption leads quite naturally to Einstein's equations themselves:

$$G^{ab} = 8\pi T^{ab}, \tag{5}$$

because G^{ab} has covariant-divergence zero as a mathematical necessity. This is why in most derivations of (5) in texts, the assumption of 'the conservation law' (1) is always made, as the crucial step that singles out the Einstein tensor G^{ab} as the correct gravitation tensor for GTR. But as we have seen, ironically, the received view at the same time concedes that (1) is not really a conservation law at all!

There is more to say than this, however. The received view concedes that (1) is not a conservation law for *material* stress-energy alone, but it instead presents it as a true conservation law for the sum of material *plus gravitational* stress-energy.

Since (3) applies only in the absence of gravity, it is natural to suppose that it might be generalised by including a term t^{ab} for *gravitational field* stress-energy thus:

$$(T^{ab} + t^{ab})_{,b} = 0. \tag{6}$$

The idea here is that, though T^{ab} alone is not conserved in the sense of equation (3) when gravity is present, perhaps a quantity representing the energy-momentum contained by the gravitational field exists such that the *sum* of the two is a conserved quantity. The existence of a non-zero divergence of this second quantity would then indicate the conversion of energy from ordinary to gravitational or *vice versa*. And, like the classical conservation law (3), this formula can be taken as generally covariant. Unfortunately, things still do not work out as smoothly as one might hope.

The object representing stress-energy of the gravitational field, t^{ab}, is defined indirectly. First, the *covariant* differential conservation law (1′) is postulated; then equating (6) and (1′) gives an implicit definition of t^{ab} in terms of the metric (*via* the components of the affine connection). But t^{ab} is not a tensor, and is not *uniquely* defined by equations (6) and (1′). It is called the gravitational stress-energy *pseudo-tensor*, and its non-tensorial nature means that there is no well-defined, intrinsic 'amount of stuff' present at any given point. In particular, unlike a genuine tensor, t^{ab} can be made to vanish at any given point by suitable coordinate transformations. Equations (6) and (1′) therefore cannot really be telling us about the *local* interchange of energy-momentum between gravity and matter.[10]

[10] Many textbooks in fact leave out this claim about interchange, conceding the non-local character of gravitational stress-energy and moving directly to the third stage (described below).

So, as with gravitational potential energy in Newtonian physics, the received view says that gravitational stress-energy is non-localisable; but it insists that it is nevertheless real, no mere bookkeeping device.

We are now ready for stage three of the received view. Having first admitted that (1) is not a genuine conservation law for material energy-momentum, and now having admitted that it also cannot be taken as describing an *intrinsic, local* interchange of energy between gravity and matter, the fallback position is:

> Still, *globally* it is true that (6) describes the interchange of energy and momentum between gravitational and ordinary stress-energy, and this is shown by the fact that $(T^{ab} + t^{ab})$ satisfies integral conservation laws. Because (6) $[(T^{ab} + t^{ab})_{,b} = 0]$ is generally covariant and involves only ordinary divergences, we can apply Gauss' theorem:
>
> $$d(P^b)/dt = d/dt \int_V (T^{0b} + t^{0b})\, d^3x = -\int_S (T^{ib} + t^{ib})\, n_i\, dS \qquad (7)$$

(Here I choose a more specific application of Gauss' theorem than in (4) above.) But because of the pseudo-tensorial nature of t^{ab}, this is misleading as well.

The pseudo-tensor nature of t^{ab} results in the integrals in (7) being ill-defined, coordinate-dependent in general. They do however yield unambiguous results, and constitute a conservation equation, when certain conditions are imposed. For a well-defined result,

(a) Integrals must be taken in limit $r \to \infty$
(b) Asymptotic flatness of the spacetime is assumed ($g^{ab} \to \eta^{ab}$)
(c) The coordinate system must be Lorentzian asymptotically (but can vary arbitrarily in the interior).

What do these integrals (7) give us? P^b is a 'Lorentz-covariant 4-vector' existing 'in' the asymptotic flat space at infinity. The equations can be applied to calculate energy loss by a system due to gravity wave transport; such calculations agree with observations on at least one binary star/pulsar pair, and this is currently taken both as proof of the existence of gravity waves, and further confirmation of GTR in general.

But whether or not this practical application of (7) is correct, it should be clear that the stringent limitations make (7) no genuine conservation principle for GTR. First and foremost, of course, we know the actual world is not asymptotically Minkowskian. So (7) does not really apply to gravity in the actual world (though, again, it may be decently approximated for some relatively isolated systems). But second, and equally importantly, holding up (7) as an important physical result goes against the most important and philosophically progressive approach to spacetime physics: that of downplaying coordinate-dependent notions and effects, and stressing the intrinsic, covariant and coordinate-independent as what is important.[11]

Although it is definitely part of the received view to put forth (7) as an important and meaningful substitute for classical conservation laws, it is also not uncommon to express open discomfort or disappointment with this result. I will close this section by citing one such expression, which points in the direction I want to advocate myself (Stephani, 1990, p. 142):

> Before turning to [the problem of finding the field equations], we want to formulate clearly the alternatives confronting us. Either we wish to calculate only with tensors and allow only covariant statements, in which case we use [(1)] and can write down no balance equation for the energy transport by radiation. Or else we want such a balance equation [(6)], which can only be formulated in a non-covariant manner; as one can see from [(6)], t^{ab} is not a tensor ...
>
> Since we pick out a Minkowski metric in the far field in a non-covariant fashion anyway, to begin with we will accept the lack of covariance in [(6)], not going into its consequences until later. There are, however, good reasons for deciding to maintain covariance and to regard the question of energy transport by gravitational waves as inappropriate in the theory of gravitation, because the concept of energy has lost its meaning there.

4. A Different Perspective

There is no genuine energy-momentum conservation principle in GTR. The above discussion makes that clear, and recall that I was essentially just laying out the standard, received view of the question as it is found in all textbooks. What typically hides this conclusion from view in these texts is the universal, almost desperate desire to make it *seem* as though there is such a principle at the heart of the theory.

I think we should abandon this effort to gloss over the facts. Let the textbooks admit openly that gravitational field stress-energy is not well-defined or fundamental, and that neither it nor ordinary stress-energy is conserved in GTR. This does not have to be understood as a defect of the theory, or a problem to be resolved (but if it is, better that it should be tackled rather than papered over!). Nor would it imply that further efforts to understand gravitational field energy and how it can be conceived to interact with matter are without value.

The perspective I advocate is a reasonable one given what the received view tells us about (1)–(7). But like any interpretation of a physical theory, it is hostage to the fortunes of new discoveries or theoretical modifications. There

[11] Weinberg (1972) says: 'The energy-momentum "vector" [P_b] defined by [(7)] is conserved, is a Lorentz four-vector, and is additive. What more could we ask?' (p. 171). The answer is obvious: we could ask for a notion of energy conservation well-defined both locally and over regions, for GTR spacetimes in general and not just one unrealistic sub-class of them.

continues, for example, to be work done on various ways of defining *quasi-local* energy and momentum of matter + gravitational field (energy-momentum contained within a finite boundary-surface), showing for example that equations governing it coincide with older asymptotic quantities such as Bondi and ADM masses. Though such proposals seem either to presuppose specific spacetime structures or have other undesirable properties, it cannot be ruled out that some such quantity emerges as satisfactory and puts gravitational stress-energy onto firmer foundations.[12]

5. Energy-Momentum and the Substantial/Relational Debate

Let us return to the relevance of the question of energy-momentum conservation to philosophical disputes about how to interpret spacetime in GTR. The clearest lesson to be drawn is that, at present, it is by no means clear that empty space must be thought to contain a genuine form of energy-momentum. The main reason for thinking there *must* be such energy is the intuition that some global energy conservation must hold in GTR, and a conviction (now supported by binary star observations) that isolated material systems such as binary star systems do lose energy over time. The energy, one naturally supposes, is being 'carried off' in gravitational form. But this intuition is undermined by an honest appraisal of the candidate general conservation principles in GTR ((1), (6) and (7)). If energy is not conserved quite generally, there is no need to make up a story about where it has gone when a system loses it.

If empty spacetime need not be thought to possess genuine energy, at least one reason for considering it to be a substance is deflected. What about the effects of gravity waves—the knocking down of walls and lighting of light-bulbs that is supposed to be possible in principle? As I said above, it is possible to view energy loss by a gravity wave source as genuine loss, without insisting that the energy is still around somewhere. (The point that emerges from the received view is that we cannot say anything about *where* such energy has gone to, even if we wanted to.) Similarly, energy gain in a gravity wave detector could be thought of as genuine gain, without our having to say that the energy existed somewhere beforehand. Such a perspective seems to strain our general cause–effect intuitions by positing a cause–effect relationship without an intermediary carrier. But the intermediary—the metric of spacetime g^{ab} with its waves—does not have to be rejected *tout court*. Relationists in any case need to give an account of how matter and its relations gives rise to spacetime structure.[13] The new challenge

[12] Quasi-local energy definitions were first proposed in the 1960s by Hawking, Penrose and others. See Hayward (1994) for a recent proposal and review of the issue.

[13] I have in mind here the strong form of relationism, called 'reductive' relationism by some authors, which sticks close to the spirit of Leibniz and Mach's views, insisting that spacetime structure is not primitive, but rather arises as a result of fundamental properties and relations of matter.

that gravity waves, or gravitational field stress-energy generally, seemed to pose was that if empty spacetime possessed energy content as real as ordinary material energy, it would be impossible to deny it an ontological status equivalent to that of ordinary matter, as substantivalism urges. This new argument for substantivalism fails as long as this alleged energy content remains ill-defined both locally and globally.

Work on quasi-local energy may eventually result in a consensus about how to define the energy content in a region of matter-free spacetime, in a way that revives the argument for substantivalism just described. But until such a time, relationists can concentrate on their traditional difficulty—explaining metrical and inertial structure without presupposing spacetime in a substantial sense.

Finally, I want to return to equations (1) and (7), and their relationship to Einstein's Machian goals in his work toward GTR.[14] Consider the integral, global energy conservation of (7). The conditions needed for deriving that result are basically this: spacetime is almost perfectly flat, almost perfectly like the absolute spaces of Newton and Minkowski; and we look at the material system from far away. As Noether's theorems showed, conservation principles emerge when we have symmetries. From a large distance, the island system of (6) is essentially just the same as a small body in a perfectly flat STR spacetime, obeying the same trivial, symmetric dynamics. That is why a Lorentz-covariant momentum vector can be assigned, and a balance equation derived. But thinking of such a physical situation as paradigmatic, or ideal in some way, is completely antithetical to Mach's Principle. For Mach's idea was that familiar Newtonian dynamics should hold in a universe full of matter everywhere; but in a near-empty spacetime—if such is even possible—things might be radically different. Throughout the period 1913–1916, Einstein's Machian goals co-existed with a contradictory insistence on finding a theory that would yield Minkowski spacetime as the ideal matter-free solution. The received view's tendency to look for something like (7) to emerge from a proper understanding of gravitational field energy is of a piece with this anti-Machian approach.[15]

A similar story may apply in the case of the covariant divergence law (1). Recall that insisting that (1) hold can be thought of as a consequence of insisting that *gravitational* physics involve 'minimal coupling' between spacetime and matter. As I argued above, Einstein's equations themselves should be seen as a sort of curvature coupling of matter to spacetime; but because of (1), that coupling is as minimal as can be in the context. If material stress-energy is not

[14] See Hoefer (1994) for an exposition of the history of Einstein's work toward GTR which shows that his main intention throughout was to cash out Mach's ideas on inertia in a workable gravity theory.

[15] By 1917 Einstein was clearly aware that his earlier insistence was antithetical to Mach's Principle, and this was the chief motivation for his introduction of the cosmological constant: he hoped to eliminate the possibility of matter-free spacetimes in the theory. (See Hoefer (1994) for details.) The discussion of this section may be seen as deepening the puzzle of Einstein's inadvertent anti-Machian assumptions.

really conserved in the classical sense (2), nevertheless we have no unusual gravity-linked behaviours of the kind one might imagine. Fluid dynamics is as much like it was in STR as can be, given that spacetime is to be curved. And again, a near-empty spacetime is the setting in which it *most closely* resembles classical behaviour; whereas Mach's Principle would tell you to look for this correspondence only in a world uniformly full of matter.

What sort of alternative behaviours or effects might one imagine? Ironically, it seems that Einstein himself was hoping to find one in his new theory. One of the main 'Machian effects' that Einstein initially thought should follow from his theory was one that he thought followed intuitively from Mach's Principle: the inertia of a test body should be increased, if we pile up a great amount of other mass in its immediate neighbourhood.[16] This could only mean an increase in inertial mass for the test body, without any overt interchange of energy-momentum explaining the increase. The minimal coupling enforced by (1), I suggest, rules out any such effect.[17]

The insistence on (1), then, may also be a way in which Einstein steered himself away from his own Machian goals without realising it. I will end by sketching a view of Einstein's equivalence principle, and where it can lead, that supplements these considerations about energy conservation and Mach's Principle.

Einstein's equivalence principle (EEP), as is well known, was a key step leading him to the idea of curved spacetime and the final theory. The EEP differs from the SEP and Weak EP normally discussed in textbooks, as John Norton (1985) convincingly showed. Einstein's principle equated the experiences of an observer in a uniformly accelerated frame with those of an observer 'at rest' in a uniform gravitational field; it was not restricted to the origin or a small neighbourhood in the region of a local inertial frame, and it did not apply to gravitational fields generally, only to the (physically unrealistic) case of a uniform field. But the significance of EEP for Einstein was profound, because (a) it signified an extension of the relativity of motion to uniform accelerations, hence further chipping away at the concept of *absolute* motion that he detested; and (b) it led naturally to the idea of thinking of the inertial structure of space, even far from matter (e.g. between galaxies) as being *itself a kind of gravitational field*. It led, in other words, almost straight to Mach's ideas on inertia: a traveller in an accelerating spaceship, out in the space between galaxies, may consider herself as being *at rest*, but in the sway of a particular gravitational field caused by all the other matter in the universe and its motion relative to her.

There are two roads one can take from EEP toward a new theory of gravity. One is the road to Mach's Principle just sketched. The goal would be to try to

[16] Barbour (1992) disputes that such an effect really follows from Mach's idea of the origin of inertia.

[17] In a similar vein some physicists have claimed that an 'anisotropy of inertia' is derivable from Mach's ideas. For example, the inertial mass of a test body near a neutron star should appear different for attempts to accelerate the mass in a direction perpendicular to the line connecting the two than it appears for accelerations along that line. Such effects do not occur in GTR, since it obeys the Very-SEP.

generalise that Machian perspective on EEP, which Einstein had successfully built into a theory by 1911, to all forms of accelerated motion. Such an Extended EEP would say that all observers, in any state of motion, could consider themselves as 'at rest', with the results of mechanical experiments in her frame understandable as the natural result of the gravitational field caused by the other matter in the universe and its motions. This equivalence would be unrestricted (i.e. not limited to a point or small region).

The other road is the one Einstein actually (and in a sense inadvertently) took, despite his Machian ambitions. Instead of generalising EEP, one can acknowledge its unrealistic aspect (no real gravity fields are homogeneous, it is at best an approximation valid in a small region). One can then reformulate the principle as a 'local' one, and focus on the case of free-fall rather than acceleration: in a LIF all physical experiments should give the same results as we predict from STR (SEP). This is still a powerful assumption, and clearly in good agreement with experience; and it leads naturally to Equation (1). But *prima facie* it has nothing to do with Mach's Principle. Instead, it constrains spacetime's structure to being Minkowskian in any LIF, without imposing any further requirements on how that structure arises.[18]

References

Barbour, J. (1992) 'Einstein and Mach's Principle', in J. Eisenstaedt and A. J. Kox (eds), *Einstein Studies. Vol. 3: Studies in the History of General Relativity* (Boston: Birkhäuser).

Callender, C. and Huggett, N. (2000) *Physics Meets Philosophy at the Planck Scale* (Cambridge: Cambridge University Press).

Cattani, C. and De Maria, M. (1989) 'Conservation Laws and Gravitational Waves in General Relativity (1915–1918)', in J. Stachel and D. Howard (eds), *Einstein Studies. Vol. 1: Einstein and the History of General Relativity* (Boston: Birkhäuser).

Einstein, A. (1916) 'The Foundation of the General Theory of Relativity', originally in *Annalen der Physik*; reprinted in translation in H. A. Lorentz, *The Principle of Relativity* (New York: Dover, 1924), pp. 109–164. (All page references are to the Dover edition).

Hayward, S. (1994) 'Quasi-Local Gravitational Energy', *Physical Review* **D49,** 831.

Hoefer, C. (1994) 'Einstein's Struggle for a Machian Gravitation Theory', *Studies in History and Philosophy of Science* **25,** 287–335.

Norton, J. (1985) 'What was Einstein's Principles of Equivalence', *Studies in History and Philosophy of Science* **16**, 203–246.

Rindler, W. (1977) *Essential Relativity* (Berlin: Springer-Verlag).

Schutz, B. (1985) *A First Course in General Relativity* (Cambridge: Cambridge University Press).

Stephani, H. (1990) *General Relativity: An Introduction to the Theory of the Gravitational Field, 2nd edn* (Cambridge: Cambridge University Press).

Weinberg, S. (1972) *Gravitation and Cosmology* (New York: John Wiley).

[18] This interpretive road was not taken by Einstein alone, but rather emerged as the physics community's understanding of the equivalence principle over time. Einstein himself tended to reserve the term 'equivalence principle' for his original EEP when explaining the conceptual foundations of GTR (see Norton, 1985).