Particles, space and time†

Vincent Icke
Sterrewacht Leiden and
Universiteit van Amsterdam

1. Clash of titans

When an extraterrestrial or a child asks me what we have *really* learned about the nature of the Universe, I reply: *our Universe consists of particles, space and time.* Even for adult terrestrials, this does not seem to be too bad a summary, so I will use it as a starting point for a brief overview. I will try to show that this seemingly innocuous statement harbours a grave problem, in the following sense: the theory of particles (called somewhat preemptively the *Standard Model*) and the theory of space and time (General Relativity, GRT) are extremely good in their own realm of application, the world of the very small and the very large, respectively. Taken together, however, we get a theory that does not conform in the least to the world as we know it.

This shows up most clearly in the fact that the Standard Model predicts that the energy density of the vacuum is not zero, in the sense that vacuum fluctuations are always present in it. Oversimplifying somewhat we may say that in relativistic quantum mechanics the uncertainty relations imply that one can never be quite sure that a cubic metre of spacetime is empty. Particle number is no longer a constant of the motion.

These vacuum fluctuations contribute to the effective mass density of the Universe. Suppose that we have a field with quanta of mass m. Then integration over momentum space shows that the energy density e_V of the vacuum is

$$e_V = (2\pi)^{-3}\hbar^{-3}\int_0^K 4\pi k^2 \frac{1}{2}\sqrt{k^2c^2+m^2c^4}\;dk \simeq \frac{cK^4}{16\pi^2\hbar^3} \quad \mathrm{J\,m^{-3}} \tag{1.1}$$

in which $K \gg m$ is the upper cutoff of the wave number. By itself this need not be a meaningful number; we could transform it away by a renormalization procedure. However, in GRT this cannot be done and we must consider the gravitational effects of this vacuum energy density. In the case of GRT we expect that K is inversely proportional to the coupling constant $8\pi G/c^2$, and so

$$K \approx \sqrt{\frac{\hbar c^3}{8\pi G}} \tag{1.2}$$

$$e_V = \frac{c^7}{1024\pi^3\hbar G^2} = 1.46\times 10^{109} \quad \mathrm{J\,m^{-3}} \tag{1.3}$$

† Invited review presented at the joint meeting of the University of Arizona and Pima Community College, *Models of Time in Physics and Cosmology*, Tucson, Arizona, 11-13 April 1996.

Considering that the energy density equivalent in an Einstein-De Sitter universe would be

$$e_0 = \rho_0 c^2 = \frac{3H_0^2 c^2}{8\pi G} = 4.24 \times 10^{-10} \quad \mathrm{J\,m^{-3}} \tag{1.4}$$

for a Hubble parameter of $H_0 = 50$ km s^{-1} Mpc^{-1} (which is 1.62×10^{-18} s^{-1}) we conclude that the clash between Einstein on the one hand and Dirac and Feynman and their cohorts on the other, is a clash of titans indeed: they differ by a factor of 10^{118}! Even if we only took K from quantum electrodynamics we would expect that it would be roughly equal to $m_e c$, and we would still get

$$e_V = \frac{m_e^4 c^5}{16\pi^2 \hbar^3} = 9.00 \times 10^{21} \quad \mathrm{J\,m^{-3}} \tag{1.5}$$

still at least a factor 10^{31} larger than observations appear to allow. It is manifestly false that $\Lambda = 10^{118}$. If it isn't anywhere near, it might as well be zero. *Why doesn't space weigh anything?*

Such a massive pile of problems is a cairn along the winding path to a region where good new physics can be found. Nobody has yet returned alive from that land; let us see what difficulties we expect to find there.

2. Spacetime as real stuff

First, consider space and time. Ever since Descartes we have known that true emptiness cannot exist; he conjectured that a force comes about through the direct physical contact between objects. In the second part of his *Principes de la Philosophie*, Article 16, Descartes wrote:

> *Concerning emptiness, in the sense given to that word by the philosophers, namely a space containing no substance, it is obvious that there is no such space in the universe, because the extent of space around or enclosed by an object is not different from the extent of that object. And even as from the sole fact that an object is extended in length, height and depth, we have reason to conclude that it is a substance, because we suppose that it is not possible for anything to have no extent, we must conclude the same about space which is supposedly empty: namely that, because it possesses spatial extent, it also has substance.*

In other words: space has physical attributes, namely its three dimensions, so it must be regarded as real stuff. This powerful notion lay hidden for three hundred years, until it was rediscovered independently by Einstein.

3. From global to local Lorentz symmetry

Ever since Einstein we have known that space and time are part of the stuff of our world, and not invisible graph paper. Thus, efforts to determine the structure of particles go in parallel with the search for the structure of spacetime. Einstein gave us a geometrical answer for the latter: a distance recipe (Lorentz-Minkowski) suffices. The theory boils down to the patching together of local Lorentz frames into a global whole, which gives it the form of a gauge field theory based on local Lorentz symmetry.

Special relativity starts from the invariance of the speed of light. The equation of motion for light is the equation for a sphere in 3-space: $s^2 = c^2t^2 - r^2$ where $s = 0$ for a light ray. Now we note that the above leads to a maximum value $v \leq c$ for all speeds v. That means that the above global Lorentz symmetry cannot be maintained. To stay consistent, we may require *local* symmetry only. After all, if signals travel with a finite speed, how would our colleagues at Arcturus know that we have just performed some Lorentz transformation here? We must restrict ourselves to the infinitesimal patch around the origin of our arbitrarily chosen standpoint:

$$ds^2 = c^2dt^2 - dr^2 \tag{3.1}$$

Local Lorentz symmetry means that, wherever you are, you can always find coordinates such that the above holds ('freely falling coordinates'). Thus, your neighbours in spacetime will also be able to do this. However, it isn't guaranteed that you will agree with the neighbours that their coordinate system $\{x^\mu\}$ is the same as yours. The best you can hope for is a patch-up between the two of you, by means of a bilinear form in the infinitesimal coordinates of some common coordinate system $\{\xi^\mu\}$ which you've both agreed to use, and for which local Lorentz symmetry holds too:

$$ds^2 = g_{\mu\nu}d\xi^\mu d\xi^\nu \tag{3.2}$$

The equation of motion corresponding to the spacetime structure encoded in $g_{\mu\nu}$ can be found by an argument similar to what one uses in classical mechanics. First, consider free motion only, the kinematics of motion without force. Classical motion is subject to certain symmetries, namely Galilei symmetry and time reversal invariance. These symmetries generate conserved quantities: momentum and energy. That is to say, the state of rest is equivalent to motion with constant energy and momentum: $dv^i/dt = d^2x^i/dt^2 = 0$. Deviations from inertial motion are attributed to an external force, $dv^i/dt = F/m$. This was basically the method that Galilei and Huygens used in their quest for the equations of motion of classical mechanics.

In special relativity we use Lorentz symmetry instead of Galilei; in GRT we extend this to local Lorentz symmetry and find that the structure of spacetime is given by $g_{\mu\nu}$. Thus the equivalent of motion under the influence of a force is force-free motion in curved spacetime. Loosely speaking: "curved spacetime gives curved paths".

Now let us proceed to the algebraical expression of that statement. A classical free particle moves according to the law of inertia: $dv^i/dt = d^2x^i/dt^2 = 0$. If an external force is present, say due to a gravitational potential Φ, we have $d^2x^i/dt^2 = -\partial\Phi/\partial x_i$. In GRT the analogue of classical free motion must be written as $du^\mu/ds = 0$, where $u^\mu \equiv dx^\mu/ds$ is the four-velocity, having changed from using the time derivative to a derivation with respect to the interval s. If we arrange with 'the neighbours' elsewhere in the Universe to refer all descriptions to a global but otherwise arbitrary coordinate system $\{\xi^\mu\}$ (much as at a conference one usually agrees to speak English), we find

$$u^\mu = \frac{dx^\mu}{ds} = \frac{\partial x^\mu}{\partial \xi^\alpha}\frac{d\xi^\alpha}{ds} = v^\alpha \frac{\partial x^\mu}{\partial \xi^\alpha} \tag{3.3}$$

Here v^α is the *local* four-velocity, the 'free-fall motion'. This produces an expression for the way in which the four-velocity changes with respect to interval:

$$0 = \frac{du^\alpha}{ds} = \frac{d^2\xi^\alpha}{ds^2} = \frac{d}{ds}\left(\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{dx^\mu}{ds}\right) \tag{3.4}$$

and, after working out the differentiations, one uses the metric tensor to obtain

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \tag{3.5}$$

$$\Gamma^\lambda_{\mu\nu} \equiv \frac{\partial x^\lambda}{\partial \xi^\alpha}\frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \tag{3.6}$$

The above equation of motion can be interpreted as "curved spacetime gives curved orbits": the four-acceleration is no longer zero but proportional to Γ, in which we have dumped all the garbage due to the 'mismatch' in spacetime caused by the introduction of local Lorentz symmetry. Thus, although we can always *locally* transform away the curvature of spacetime ("freely falling coordinates"), we cannot do so globally.

4. Space, time and particles connected

Einstein made a conjecture that encompasses the slogan 'Nature is made of particles, space and time' in a particular way. He did this by using, on the left hand side of his famous field equation, non-Euclidean geometry to describe the structure of spacetime via a local generalization of the global Minkowski distance recipe. On the right hand side he used a description of matter based on two approximations. First, he used a continuum representation of matter, averaging over the individual particles to obtain densities. Second, he connected the presence of matter to the structure of spacetime by demanding correspondence with the Newtonian equations in the limit for weak fields and small velocities. Accordingly, his equation reads something like

$$\left\{\begin{matrix}\text{structure of}\\ \text{space and time}\end{matrix}\right\} = \left\{\begin{matrix}\text{distribution of}\\ \text{particles and fields}\end{matrix}\right\}$$

If we read the equals sign as implying an interaction, this says 'particles interact with spacetime'. Of course this interpretation immediately shows that the equation is incomplete, because the back-reaction of the structure of spacetime on the properties of the particles is not included, other than by their trajectories as pointlike test particles in the field $g_{\mu\nu}$.

How are matter and spacetime curvature connected algebraically? In other words, what connects matter and the distance recipe in spacetime? When we compare the Newtonian and the Einsteinian equations for the trajectory of a particle, we notice immediately that Φ and $\Gamma^{\lambda}_{\mu\nu}$ are apparently related. Of course it cannot be that $\Gamma = \Phi$ or something as simple as that, because then Newtonian theory would be relativistic already! We must incorporate a physical ingredient in our theory. The one that Einstein used is: in the limit for small velocities and small curvatures in a static spacetime we must recover the classical equations.

In that case, the only components that remain are those related to the time-indices: everything is zero except Γ^{μ}_{00}. Using the above calculations one may see that this is closely related to the fact that the limit of the four-velocity u^{μ} for small three-velocities is $u^{\mu} = (c, 0, 0, 0)$. The zero-component does not vanish, just like the 0-component of the energy-momentum does not vanish for $v \to 0$ (this is the famous mc^2-term). Furthermore, from the definition of the metric we know that $ds^2 \simeq g_{00}dt^2$ (the other components of $g_{\mu\nu}$ vanish in the limit for small curvature), and therefore the low-velocity limit of Eq.(3.5) is

$$\frac{d^2x^{\mu}}{dt^2} + \Gamma^{\mu}_{00} = 0 \tag{4.1}$$

from which we finally conclude that the desired correspondence with classical gravitation can be obtained if we relate Γ to the gradient of the gravitational potential:

$$\Gamma^{\mu}_{00} \Longleftrightarrow \frac{\partial\Phi}{\partial x_{\mu}} \tag{4.2}$$

In order to connect with the classical Newtonian case, we felt obliged to equate Γ with a coordinate derivative of the classical potential. Then Γ is related to a linear combination of first derivatives of $g_{\mu\nu}$:

$$\Gamma^{\kappa}_{\lambda\mu} = \frac{1}{2}g^{\nu\kappa}\left\{\frac{\partial g_{\mu\nu}}{\partial x^{\lambda}} + \frac{\partial g_{\lambda\nu}}{\partial x^{\mu}} - \frac{\partial g_{\mu\lambda}}{\partial x^{\nu}}\right\} \tag{4.3}$$

and we finally conclude that *the role which the potential Φ plays in classical Newtonian gravity is taken over by the metric tensor $g_{\mu\nu}$ in general relativity.*

The classical potential is related to the presence of matter by means of the Poisson equation

$$\Delta\Phi = 4\pi G\rho \tag{4.4}$$

which should give us a hint about how the presence of matter can be connected to the structure of spacetime. The point here is that the density field ρ could not possibly be used, because it is in no way Lorentz invariant. In fact, we can see immediately that a Lorentz transformation of ρ should go as

$$\rho' \propto \gamma^2\rho \tag{4.5}$$

One Lorentz factor γ comes from the change of the effective mass via $E = \gamma mc^2$; the other one comes from the fact that a volume seen in motion decreases by one factor γ because of its Lorentz-FitzGerald contraction. Accordingly we suspect that ρ should be part of a tensor because a Lorentz scalar transforms with γ^0, a vector with γ^1 and a tensor with γ^2. Since ρ is classically a scalar field we also suspect that it is the 00-component of a tensor.

In the limit of small velocities without external forces we may take that tensor, which we'll call $T_{\mu\nu}$, to be diagonal. Then ρ is placed in the top-left corner. What will we have on the remaining three places of the diagonal? Since ρ is a mass density, and since in relativity we have to take mass and energy as equivalent, it seems natural to use an energy density. The mass density is the mass of a collection of particles per unit volume. The corresponding energy per unit volume we know as the *pressure* P of the collection of particles. The entries elsewhere in the tensor can be found by Lorentz transformation of T; the complete form is the *energy-momentum tensor*

$$T_{\mu\nu} = P\eta_{\mu\nu} + (\rho c^2 + P)u_\mu u_\nu \tag{4.6}$$

The final order of business now is to construct a tensor $G_{\mu\nu}$ from $g_{\mu\nu}$ and its derivatives that has the same transformation properties as $T_{\mu\nu}$. The most obvious choice, taking T simply proportional to g, is not sufficient because it would not include Newtonian gravity; in order to obtain that, as we had seen above, we must include the derivatives of g. In particular, in the Newtonian limit we retain the 00-components only, namely

$$\frac{\partial^2 g_{00}}{\partial x^\alpha \partial x_\alpha} \propto T_{00} \tag{4.7}$$

Apparently, the desired tensor must contain at least second derivatives of g. This implies that *a fourth-rank tensor must be involved!* Einstein showed that the correct expression is related to the monstrous *Riemann-Christoffel curvature tensor*

$$R^\lambda_{\mu\nu\kappa} \equiv \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial x^\kappa} - \frac{\partial \Gamma^\lambda_{\mu\kappa}}{\partial x^\nu} + \Gamma^\alpha_{\mu\nu}\Gamma^\lambda_{\kappa\alpha} - \Gamma^\beta_{\mu\kappa}\Gamma^\lambda_{\nu\beta} \tag{4.8}$$

This animal must be reduced to second rank before we can equate it to $T_{\mu\nu}$. This is done by contracting it over one index. The most general expression for the required tensor is then

$$R^\lambda_{\mu\lambda\nu} - \frac{1}{2}g_{\mu\nu}g^{\alpha\beta}R_{\alpha\beta} + \Lambda g_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} \tag{4.9}$$

and in units where the gravitational constant G is retained explicitly, correspondence with the Poisson equation shows that

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^2}T_{\mu\nu} \tag{4.10}$$

This is a physical choice; it does not have the mathematical necessity of a gauge theory because of the way in which $T_{\mu\nu}$ was put in by hand. It is Einstein's guess, based on correspondence with Newtonian mechanics, and it works very well on large scales: black holes, relativistic stars, the Universe. But it works not at all on small, atomic, quantum-mechanical scales. The hassle is that we are obliged to include the vacuum zero-point energy in $T_{\mu\nu}$. As we saw in Sec.1, it shows up in the form of a finite value $\Lambda \approx 10^{118}\rho_0 c^2$of the cosmological constant which is totally excluded by cosmological observations.

Possibly there is another guess one could make by searching for an expression that corresponds with (say) Schrödinger's Equation instead of Poisson's Equation, but nobody has yet succeeded in doing so.

Note that the rather peculiar and counter-intuitive behaviour of Λ, and the related possibility of inflation ("you get something out of nothing") is due to our initial assumption for the connection between $g_{\mu\nu}$ and the potential Φ, and via Φ and the Poisson equation to the (thermo)dynamic mass- and energy densities ρ and P. The connection between matter and spacetime curvature is still a conjecture, since we do not have a quantum gravity theory.

One immediate cause for worry is that it seems like 'double dipping' to introduce spacetime in $g_{\mu\nu}$ as well as in Λ. Ought we not to include the Planck-scale fluctuations in some (possibly extended) form of $g_{\mu\nu}$ rather than in the potential term that produces Λ? After all, a potential is a classical beast that would be wiped out by second quantization.

5. Twists and wrinkles

Having described the structure and behaviour of space and time as they appear in our summary expression 'the Universe is made of particles, space and time', let us consider the particles. The trick of getting a field $g_{\mu\nu}$ that corrects for the consequences of using a *local* Lorentz symmetry instead of a global one is common to all current theories of interaction. First let me try to explain in somewhat pedestrian terms what happens here.

The similarity between the symmetries of Nature and simple rotations enables us to understand how a symmetry can produce a force field. Take before you, on a smooth table, a small tablecloth, or something similar (e.g. a large piece of aluminium foil). The material must be a uniform colour, without any patterns. Make sure it is quite smooth. We are not looking so closely that the individual fibres are visible, and we will pretend that the material extends to infinity: our tablecloth is a small piece of an unbounded model universe. Now rotate the whole cloth through an arbitrary angle. Any piece of the surface, when inspected individually, appears the same as before: the cloth is *invariant under global rotations.*

But it would be impossible, even in principle, to do something like this with the real Universe. Imagine that we want to perform a global symmetry transformation. Then we would have to let the symmetry act in all of space at exactly the same time. But this is impossible to do in reality because no signal can propagate faster than the speed of light. We must *accept only local symmetry rotations*, that is, a symmetry where the amount of rotation *differs* from event to event in spacetime.

Return to the tablecloth before you. Put your finger on a point near the centre and give the cloth an arbitrary twist, keeping the edges of the cloth in place. When you remove your finger, you notice that the piece of the surface you have just rotated still appears the same; at that one point, the cloth is invariant under local rotations. But in the vicinity of the twisted point, something has happened: *a spray of wrinkles radiates outward from it.* The local twist cannot be connected smoothly with the undisturbed cloth at large distances: the difference must be patched up. Because of relativity, all symmetries must be local, and *any local symmetry creates a field.* The wrinkles are related to the 'field lines'.

6. Gauge twists and velpons

Elementary particles belong to certain families. Within such a family the particles treat each other as equals, at least in the ideal case. That is to say that they are, in some sense, interchangeable: because of the perfect equality one would not notice such a swap. This operation is a symmetry. If one were to subject the whole Universe at once to such a *global* symmetry, nothing would change at all.

If we pick a fundamental fermion multiplet with N members we expect that the symmetry group that acts on this N-plet should behave as a rotation in some abstract N-dimensional space. The rotation of a particle over the *mixing angle* θ literally makes the particle 'turn into' a different one! Some symmetries are actually connected with rotation in space (which creates the angular momentum of a particle) or rotation in spacetime (Lorentz symmetry). Other symmetries behave like rotations too, but not in ordinary space; rather, these symmetries are rotations about other directions than the axes of space and time. Apparently, the vacuum possesses more possible directions than those of spacetime.

A symmetry can be responsible for generating a field. Now in the quantum picture a field is built up from field quanta; and the exchange of a quantum produces a force. Thus, *any local symmetry creates a force.* The quanta of such a field are gauge quanta, or, more precisely, *gauge bosons*, because Lorentz invariance and exact gauge symmetry combined demand that the particle have mass zero and integral spin. The force that corresponds to the exchange of gauge bosons can be considered as a binding agent, a kind of glue between the particles to which the bosons are coupled.

Each force has its own set of glue quanta. A generic name (other than the insipid 'gauge boson') for these does not exist in the professional literature. Therefore, I will succumb to the temptation to name something, and use the generic name *velpon*, after one of the most common brands of glue in my home country.

All known forces are due to gauge symmetries. A local gauge twist causes wrinkles in the vacuum, because the mismatch between the twisted space and the unperturbed vacuum in the distance must be patched up somewhere in between. In the non-quantum picture, of which our tablecloth is a model, the difference is made good by wrinkles, 'field lines' that radiate from the twisted spot. But in a quantum world, where interaction is all-or-nothing, *the twist must be taken away by a single quantum.* The exchanged quantum, the velpon, then becomes the carrier of the vacuum wrinkle caused by the local gauge twist.

7. The Lagrangian in QED and in gravity

The above explanation is only an analogy and needs to be made more precise for practical purposes. This I will do by briefly showing how the 'wrinkle' picture of fields can be cast in algebraic form. It all begins by guessing a proper symmetry from some notion of similarity between particles, in particular fermions. For example, one may note that a proton is really rather like a neutron, with the exception of a small difference in mass and a difference in electric charge. It takes a little faith to overlook these differences, but one may with some justification surmise that the mass difference will ultimately appear to be due to the charge or something like that. Or one may note that in a sense an electron is rather like a neutrino, in that they always appear together in weak decays.

Start with such a guessed global symmetry. Then one notes that using this symmetry globally is contrary to the spirit of relativity. But using a *local* symmetry is not possible unless one inserts a new field to counteract the mismatch caused by locality. The quanta of this mismatch field transmit a force. In this way, a local symmetry of a basic multiplet of fermions produces bosons that couple to the fermions in a way that is prescribed by the symmetry.

Suppose that our mechanical system is described by a Lagrangian $\mathcal{L}$, which is a function over spacetime $\{x_\mu\}$ of a generalized coordinate vector q and its corresponding momentum $q_{,\mu}$ (we use the abbreviation $q_{,\mu} \equiv \partial q/\partial x^\mu$). The action corresponding to this is found by integrating the Lagrangian density over an arbitrary four-volume Ω:

$$\mathcal{S} = \int_\Omega \mathcal{L}(q, q_{,\mu}) dx_\nu \tag{7.1}$$

where the dynamical variables q and $q_{,\mu}$ are to be seen as functions of x_μ. Because Ω is arbitrary, the requirement $\delta\mathcal{S} = 0$ implies

$$\int_\Omega \frac{\partial\mathcal{L}}{\partial q}\delta q + \frac{\partial\mathcal{L}}{\partial q_{,\mu}}\delta q_{,\mu}\, dx_\nu = \int_\Omega \left(\frac{\partial\mathcal{L}}{\partial q} - \frac{\partial}{\partial x^\mu}\frac{\partial\mathcal{L}}{\partial q_{,\mu}}\right)\delta q\, dx_\nu + \oint_\Omega \frac{\partial\mathcal{L}}{\partial q_{,\mu}}\delta q\, dx_\mu = 0 \tag{7.2}$$

from which the Lagrangian equations of motion follow directly because the surface integral is zero. Note that we are allowed to subject $\mathcal{L}$ to the same symmetry under which q is supposed to be symmetric, because $\mathcal{L} = f(q, q_{,\mu})$.

Now let us change q infinitesimally by some symmetry $\mathbf{L}$. A global symmetry does not change the equations of motion, because $\mathbf{L}$ commutes with δ. However, if $\mathbf{L}$ is a *local* symmetry, then $\mathbf{L} = \mathbf{L}(x_\mu)$, and therefore

$$q \to q + \delta q \quad \text{and} \quad \delta q = \epsilon(x_\mu)q \tag{7.3}$$

It follows immediately that if $q_{,\mu} \to q_{,\mu} + \delta q_{,\mu}$, we get

$$\delta q_{,\mu} = \epsilon_{,\mu} q + \epsilon(x_\mu) q_{,\mu} \tag{7.4}$$

so that the integrand of $\delta\mathcal{S}$ becomes

$$\delta\mathcal{L} = \left(\frac{\partial\mathcal{L}}{\partial q}q + \frac{\partial\mathcal{L}}{\partial q_{,\mu}}q_{,\mu}\right)\epsilon + \frac{\partial\mathcal{L}}{\partial q_{,\mu}}\epsilon_{,\mu}q \tag{7.5}$$

The term in brackets drops out because of the equations of motion, and we conclude that $\delta\mathcal{L} \neq 0$ because of the derivative $\epsilon_{,\mu}$: the *local* character of the transformation **L** spoils the proper extremum behaviour of $\mathcal{L}$, and no good equations of motion result!

In other words, the fact that **L** changes from event to event in spacetime produces a *mismatch* between $\mathbf{L}\mathcal{L}$ at one event and the $\mathbf{L}\mathcal{L}$ elsewhere. The key idea now is, to patch this up by adding extra terms to the Lagrangian to correct the mismatch. It is by no means obvious that this can be done successfully!

Because the culprit is a vector $\epsilon_{,\mu}$, we try to patch up $\mathcal{L}$ by adding a vector field to it. For the moment, let us call this field A', and the corresponding Lagrangian is

$$\mathcal{L}' = \mathcal{L}'(q, q_{,\mu}, A') \tag{7.6}$$

of which we will now rigorously require that $\delta\mathcal{L}' = 0$. This requirement prescribes a functional dependence of $\mathcal{L}'$ on its arguments, as follows. The δq and $\delta q_{,\mu}$ are found as before; the variation $\delta A'$ is, of course, a linear combination of ϵ and $\epsilon_{,\mu}$ (for infinitesimal transformations). The most general form for $\delta A'$ is then

$$\delta A' = UA'\epsilon + C^{\mu}\epsilon_{,\mu} \tag{7.7}$$

with constant scalar U and vector C^{μ}, to be determined afterwards. To get a proper equation of motion from the patched-up Lagrangian, we require

$$\delta\mathcal{L}' = \frac{\partial\mathcal{L}'}{\partial q}\delta q + \frac{\partial\mathcal{L}'}{\partial q_{,\mu}}\delta q_{,\mu} + \frac{\partial\mathcal{L}'}{\partial A'}\delta A' = 0 \tag{7.8}$$

Inserting the expressions for δq and so forth yields a linear equation in ϵ and $\epsilon_{,\mu}$. Because the magnitude of ϵ is arbitrary (provided it is infinitesimal), each coefficient of ϵ and $\epsilon_{,\mu}$ must vanish independently. This gives

$$\frac{\partial\mathcal{L}'}{\partial q}q + \frac{\partial\mathcal{L}'}{\partial q_{,\mu}}q_{,\mu} + \frac{\partial\mathcal{L}'}{\partial A'}UA' = 0 \tag{7.9}$$

$$\frac{\partial\mathcal{L}'}{\partial q_{,\mu}}q + \frac{\partial\mathcal{L}'}{\partial A'}C^{\mu} = 0 \tag{7.10}$$

with the consistency requirement that

$$C^{\mu}C_{\mu} = 1 \tag{7.11}$$

The latter means that C_{μ} has an inverse; if it did not, then some of the above equations would be linearly dependent and the system could not be solved. Now define the vector field A_{μ} as

$$A_{\mu} \equiv C_{\mu}A' \tag{7.12}$$

to find that

$$\frac{\partial \mathcal{L}'}{\partial q_{,\mu}} q + \frac{\partial \mathcal{L}'}{\partial A_\mu} = 0 \tag{7.13}$$

The remarkable thing is, that *this equation is in fact a prescription for the way in which the Lagrangian must depend on its arguments.* We see directly that it requires that the vector field A_μ, which was introduced to patch up the mismatch created by the locality of $\mathbf{L}$ (i.e. the dependency $\mathbf{L} = \mathbf{L}(x_\mu)$) occurs in $\mathcal{L}'$ *only* through the combination

$$q_{;\mu} \equiv q_{,\mu} - qA_\mu \tag{7.14}$$

the *covariant derivative* of q. The form $\mathcal{L}' = \mathcal{L}'(q, q_{,\mu}, A')$ allows us only one way to insert $q_{;\mu}$ into $\mathcal{L}'$, namely in exactly the same way as $\mathcal{L}$ depends on $q_{,\mu}$. This must be so because $q_{;\mu}$ contains a term that is linear in $q_{,\mu}$, and another term that can be made zero by letting $\mathbf{L}$ equal the identity. Thus we get

$$\mathcal{L}' = \mathcal{L}(q, q_{;\mu}) \tag{7.15}$$

and from now on we use this form.

Note that in the covariant derivative the local symmetry prescribes that q and A_μ couple by means of the product qA_μ; in quantum electrodynamics this appears in the form where the dynamical variables q and $q_{,\mu}$ are replaced by the derivative ∂ and a constant factor ie:

$$q_{,\mu} - qA_\mu \to \partial - ieA \to i\psi^*(\gamma \cdot \partial - ie\gamma \cdot A)\psi \tag{7.16}$$

which is the famous 'minimal coupling' term in the Dirac equation (the γ's are Dirac matrices).

Having now found that there is only one functional form of the Lagrangian which allows us to patch up the mismatch due to the local symmetry, it remains to determine the constants U and C^μ. First, we note that

$$\delta A_\mu = C_\mu \delta A' = C_\mu U \epsilon(x_\alpha) A' + \epsilon_{,\mu} = C_\mu C^\nu U \epsilon A_\nu + \epsilon_{,\mu} \tag{7.17}$$

Second, we recall the expressions for the variations $\delta\mathcal{L}$ and $\delta\mathcal{L}'$, which lead directly to

$$\frac{\partial \mathcal{L}'}{\partial q} = \left.\frac{\partial \mathcal{L}}{\partial q}\right|_{q_{;\mu}} - \left.\frac{\partial \mathcal{L}}{\partial q_{;\mu}}\right|_{q} A_\mu \tag{7.18}$$

$$\frac{\partial \mathcal{L}'}{\partial q_{,\mu}} = \left.\frac{\partial \mathcal{L}}{\partial q_{;\mu}}\right|_{q} \tag{7.19}$$

$$\frac{\partial \mathcal{L}'}{\partial A'} = -\left.\frac{\partial \mathcal{L}}{\partial q_{;\nu}}\right|_{q} C_\nu q \tag{7.20}$$

Inserting these into the equation resulting from $\delta\mathcal{L}' = 0$, we find

$$\left(\frac{\partial \mathcal{L}}{\partial q} q + \frac{\partial \mathcal{L}}{\partial q_{;\mu}} q_{;\mu}\right) - \frac{\partial \mathcal{L}}{\partial q_{;\nu}} qUA_\nu = 0 \tag{7.21}$$

The term in brackets vanishes because of the equation of motion for $\mathcal{L}$, and because we had $\mathcal{L}' = \mathcal{L}(q, q_{;\mu})$. It follows immediately that $U = 0$. The definition of A_μ then gives, by means of the expression for $\delta A'$, that

$$\delta A_\mu = \epsilon_{,\mu} \tag{7.22}$$

This demonstrates quite clearly how the vector field A_μ comes in because of the *local* character of the symmetry: if $\mathbf{L}$ were independent of x_α, we would have $\epsilon_{,\mu} = 0$!

One further point remains to be settled. By patching up the Lagrangian, we have let a genie out of a bottle, namely the field A_μ. We are now obliged to take this field seriously, and to identify it with an actual particle. In that case, we must of course allow A_μ to occur in the Lagrangian as a free field (i.e. as more than just an entity which couples to the q-field by qA_μ). It may be a trifle much to ask, but can the locality of $\mathbf{L}$ prescribe the form of the occurrence of this free field too?

The patch-up vector field A_μ may occur itself in the Lagrangian as a dynamical variable, together with its spacetime derivative $A_{\mu,\nu}$, in the same way that we had a dependence on the dynamical variables of the q-field. Because $\mathcal{L}$ is linear, we can insert extra terms by simple addition, so we can restrict ourselves to finding the subpart $\mathcal{L}''$ that depends on the A's only, and then add it to what we had already (note that the coupling term has already been disposed of!) We use the same variational form:

$$\delta\mathcal{L}'' = \frac{\partial\mathcal{L}''}{\partial A_\mu}\delta A_\mu + \frac{\partial\mathcal{L}''}{\partial A_{\mu,\nu}}\delta A_{\mu,\nu} \tag{7.23}$$

As usual, we insert δA_μ, and require that each coefficient of ϵ and $\epsilon_{,\mu}$ vanish independently. This yields the equations

$$\frac{\partial\mathcal{L}''}{\partial A_\mu} = 0 \tag{7.24}$$

$$\frac{\partial\mathcal{L}''}{\partial A_{\mu,\nu}} + \frac{\partial\mathcal{L}''}{\partial A_{\nu,\mu}} = 0 \tag{7.25}$$

Accordingly, we find that the patch-up field *cannot itself occur in the Lagrangian.* Consequently, the field A_μ *is not an observable;* but it *can* couple to the dynamical variable q by means of the term qA_μ. In Feynman terms: the A_μ can only occur between vertices, it is an intermediary, a *virtual particle.* The above shows that the new field can occur in $\mathcal{L}$ only through the combination

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu} \tag{7.26}$$

That is to say, *the curl of the field is an observable!* This should of course look very familiar to aficionados of Maxwell's Equations.

This completes the demonstration that the requirement of local symmetry of the Lagrangian is so severe, that not only the way in which the patch-up field A_μ couples to the q-field, but also the way in which it must occur in the Lagrangian is prescribed entirely. This almost total lack of arbitrariness is what makes the local symmetry concept so compelling.

It can be shown that the same kind of construction works for vector fields q^a (in fact, this is what the original Yang-Mills paper was all about). In that case, it can be shown that the strictness and cleanness with which the form of the Lagrangian is prescribed is due to the group structure of the symmetry.

We have four such cases in Nature:

(1) the case of a phase-rotation symmetry $U(1)$, (i.e. the multiplication with a complex scalar function as treated above), which produces electromagnetism;
(2) the case of the "isospin symmetry" $SU(2)$ (i.e. multiplication with a factor derived from a 2×2 symmetry via $\exp(\frac{1}{2}ig\tau \cdot \omega)$, where τ are Pauli matrices and ω is an arbitrary smooth function over spacetime), which produces the weak interaction;
(3) the group $SU(3)$, leading to the colour interaction;
(4) Lorentz symmetry, which gives rise to the gravitational interaction (General Relativity).

In the Yang-Mills case, the group has nonzero structure constants f^a_{bc}, and following exactly the same line of resoning one may show that the "wrinkle" or "patch-up" field A^a can occur in the Lagrangian only through the combination

$$F^a_{\mu\nu} = A^a_{\nu,\mu} - A^a_{\mu,\nu} - \frac{1}{2} f^a_{bc} \left(A^b_\mu A^c_\nu - A^b_\nu A^c_\mu \right) \tag{7.27}$$

which clearly shows the occurrence of nonlinear terms due to the non-Abelian character of the group. The range of the index a depends on the group dimension; for SU(N), it is $N^2 - 1$.

If the local symmetry is Lorentz symmetry, one may show in precisely the same way – though with much more effort – that the patch-up fields (which in gravity theory are traditionally called Γ instead if A) can occur in the Lagrangian only through the combination

$$R^\lambda_{\mu\nu\kappa} \equiv \Gamma^\lambda_{\mu\nu,\kappa} - \Gamma^\lambda_{\mu\kappa,\nu} + \Gamma^\eta_{\mu\nu}\Gamma^\lambda_{\kappa\eta} - \Gamma^\eta_{\mu\kappa}\Gamma^\lambda_{\nu\eta} \tag{7.28}$$

which is the Riemann-Christoffel tensor.

8. Clash

The remarkable fact is that both GRT and the Standard Model describe fundamental interactions by means of a gauge field. On small scales, these are quantum fields due to the symmetries U(1), SU(2) and SU(3). On large scales, the gauge field is the Christoffel object $\Gamma^\lambda_{\mu\nu}$ due to local Lorentz symmetry.

It would be great if these similarities allowed us to bring all known forces together in one formalism. Then the expression 'the Universe is made of particles, space and time' would get a truly compelling uniformity, so that the field equations would read something like

$$\left\{\begin{array}{c}\text{structure of}\\ \text{space, time and}\\ \text{particles}\end{array}\right\} = 0$$

Alas, we are disappointed. The structure of spacetime cannot be included with the density of particles if we use the Standard Model in the matter tensor. In field theory a potential is no longer something that can be freely changed by adding an arbitrary scalar term; due to the local (as opposed to global) character of the fields, a potential becomes an entity in itself, witness for example the occurrence of A_μ. In electrodynamics this is merely the vector potential, in quantum electrodynamics it stands for a real particle, the photon, which cannot be transformed away. Einstein's conjecture runs into profound trouble because the reality of potentials implies that the zero point energy of the vacuum must be included in the Einstein equation.

9. Pigs in space

So the theory of forces and matter tells us that in the vacuum, this space 'supposed to be empty', spontaneous particle-antiparticle pairs arise. And that means that this apparently empty vacuum plays an active role in Nature: it has properties that have a profound influence on the behaviour of particles and their interaction. This is apparent, among other things, in the anomalous magnetic moment of the electron and in the Casimir-Polder force, in which the zero-point energy corresponding to the spontaneous pairs exerts a measurable influence on the force between conductors.

If one introduces matter into space in this manner, things simply won't fit. We cannot have pigs in space, because their attendant vacuum fluctuations would ruin the Universe. Einstein built his theory of spacetime expressly in such a way that Newtonian mechanics was recovered in the limit for small speeds and weak fields. But Newtonian theory is all wrong on a small scale, so it would be a stupendous marvel if the Einstein equation gave the right result all the way. If only he had believed in quanta, maybe he could have forged a correspondence with the Schrödinger or Dirac equation!

There are several ways out. People have tried to cancel the vacuum fluctuations against each other. That is in itself not so bizarre: a force is the net result of a quantum sum over *all* Feynman diagrams, and counterdiagrams might be dreamed up, as in the case when the c-quark was predicted from the absence of the decay of the kaon into a pair of muons. But the cancellation would have to be so extraordinarily perfect that it is contrived in the extreme. Only unbroken supersymmetry seems to help. In these theories, there is a symmetry that connects fermions and bosons. Quite a desperate move, because bosons and fermions are as un-alike as possible! However, such un-kosher combinations can be made, using generators Q_α, four-momentum p^μ, Pauli matrices τ_μ and the anticommutation rule

$$\left\{Q_\alpha, Q_\beta^\dagger\right\} = (\tau_\mu)_{\alpha\beta}\, p^\mu \tag{9.1}$$

The vacuum is the state which has

$$Q_\alpha|0\rangle = Q_\alpha^\dagger|0\rangle = 0 \tag{9.2}$$

so that the anticommutator {} immediately produces

$$\langle 0|p^\mu|0\rangle = 0 \tag{9.3}$$

that is to say, the energy-momentum density of the vacuum is zero: $\Lambda = 0$. For this mechanism to work, each fermion should have a bosonic counterpart and vice versa. Our world doesn't look like that in the least, so we are not much further along.

Who's gotta yield, the left- or the right hand side of the Einstein equation? Both, in a sense, but the left more than the right. The right hand side, $T_{\mu\nu}$, is a continuum average and doesn't contain individual particles. That must modified: no more spacetime averages, no densities, because in a quantum formulation we should not expect space and time to be continuous in the conventional sense. Averages and derivatives $\partial/\partial x^\mu$ would lose their meaning.

The left hand side, the Einstein tensor, at first appears to be the strongest fortress because it is purely mathematical. However, its description of the structure of spacetime in terms of a distance recipe (via $g_{\mu\nu}$) can probably not be quantized. After all, if one were to interpret g in terms of a collection of particles instead of a classical field, we get a paradox: *if spacetime is made of particles, how could such a particle move through space and time?* Stated somewhat differently, the usual particle symmetries (U(1), SU(2), SU(3) and relatives) connect fermions and bosons, whereas the Lorentz symmetry says something about spacetime and not about particles. Furthermore, it is my prejudice that quantum behaviour is a much more strongly established physical effect than the Einsteinian variant of gravity. Anyone can see the Balmer series with minimal equipment. Compared with that loud-and-clear demonstration of quantization, the usual relativistic tests (Mercury, gravitational lensing) seem weak and indirect. A black hole would be a convincing thing, analogous to the hydrogen atom, but its properties are still inferred only indirectly.

So how can we make particles out of g? When matching $g_{\mu\nu}$ to quantum degrees of freedom we need not assign every component to a dynamical field. In QED, the classical vector potential A_μ reappears as the photon; in GRT we have a correspondence between the Newtonian potential Φ and $g_{\mu\nu}$. But part of the covariance with respect to g could be purely a coordinate effect, i.e. there might be one specific 'nature-given' set of coordinates where some fields vanish. The large-scale homogeneity and isotropy of the Universe is a case in point: spherical coordinates might be a preferred reference frame, in which case Lorentz covariance is no longer guaranteed. Having decided which components of g should appear as quantum degrees of freedom, we must decide how to assign these to observable particles. Quantum superposition allows us to construct states by linear combination, as in the case of the U(1)$\otimes$SU(2) unification: the photon γ and the neutral vector boson Z^0 are superpositions of the U(1) and SU(2) velpons. Compare also the Englert-Brout-Higgs mechanism for making particles massive: the degrees of freedom of the scalar particles are used to generate two extra degrees of freedom for the W and the Z, which – if they were massless – would have only two helicity states each, instead of the four of a massive particle.

One could make $\Lambda = 0$ at one specific point in time, but the expansion of the Universe would shift us away from that point and we'd be just as badly off. Similarly, if we inflate the Universe by using some sort of value K derived from grand unification or a similar theory, we should expect that today we'd still be not too far away from an $n = 1$ state in the bottom of the potential of the GUT, and thus have a substantial fraction of the GUT potential still around.

Currently I am trying to do away with the problem more radically by just stating that, by fiat, gravitons do not interact with vacuum fluctuations. The attractiveness is that this removes the need for renormalization of gravity: there'd be no more loops in the graviton propagator. But taken literally that should also exclude one-loop diagrams from the interaction between gravitons; gravity would no longer be nonlinear, against all the evidence.

At this point, I'm talking pie-in-the-sky. It is like Bohr's treatment of the hydrogen atom. Bohr knew perfectly well that an accelerated electron in an atom ought to radiate like crazy. But he pretended it doesn't, just to see what happens. But I'm not Bohr, and I don't know what the equivalent of the hydrogen atom is. The Schwarzschild black hole may fit the bill. But the effects I've calculated are nowhere near observable, and probably all false. Nobody, so far, has found an escape route.

10. *Envoi*

What I have discussed here is known and unknown. So well known that I'm quite embarrassed to talk about it, and so utterly unknown that I feel similarly embarrassed. In between we may be lucky to find fertile ground where we can practice what Medawar called *the art of the soluble.*

References

Icke, V., *The Force of Symmetry*, Cambridge University Press, Cambridge, 1995

Utiyama, R., *Invariant theoretical interpretation of interaction*, 1956 *Phys. Review* **101**, 1597-1607

Weinberg, S., *The cosmological constant problem*, 1989 *Revs. Mod. Physics* **61**, 1-23