Shift-symmetry in Einstein’s universe: Einstein’s quest for mathematical perfection

Eric Baird

Part 1 of a series. In this paper we identify three interlinked, interdependent, and mutually-supporting principles that run through Einstein’s work, and seem to have guided him through his career: the symmetry of the equations of inertial physics, the symmetry of the equations of gravitational physics, and time-symmetry.

Table of Contents

1. Introduction.................................................................................................................................................... 2
2. The philosophical basis of the current system:............................................................................................ 3
   A. Doppler symmetry........................................................................................................................................ 3
      A.1 Assumed absence of any effect of moving matter on light................................................................. 3
      A.2 The “bead-in-a-box” scenario.................................................................................................................. 4
         a) Newtonian result........................................................................................................................................ 4
         b) SR result, and “shift symmetry”.............................................................................................................. 4
      A.3 Energy conservation................................................................................................................................. 4
      A.4 Uniqueness of the SR solution regarding shift symmetry..................................................................... 5
   B. Gravitational shift symmetry..................................................................................................................... 6
      B.1 GR1916 route-independence of gravitational shifts............................................................................ 6
      B.2 GR1916 “shift symmetry”.................................................................................................................... 6
      B.3 Connection to motion shifts.................................................................................................................. 6
      B.4 GR1916 and the SR legacy.................................................................................................................... 6
      B.5 The Schwarzschild solution.................................................................................................................. 7
   C. Time-symmetry............................................................................................................................................ 8
      C.1 Eddington on time.................................................................................................................................... 8
      C.2 Uniqueness of the SR solution regarding time symmetry..................................................................... 8
3. Apparent impregnability of the SR-GR1916 system.................................................................................... 9
4. Conclusions.................................................................................................................................................... 9
References.......................................................................................................................................................... 10

1. Introduction

The search for methods of unifying and expanding theoretical physics is often made easier by the use of restrictive principles, the hope being that a restrictive principle not only encourages minimalism and limits the number of possibilities that need to be considered, but may be so restrictive that perhaps, in order to satisfy it, the number of possibilities may reduce to just a single solution.

Examples of restrictive principles include the special principle of relativity (“SPoR”), the general principle of relativity (“GPoR”) the principle of equivalence (of inertial and gravitational mass, “PoE”), the law of energy conservation (various versions), Bohr’s “correspondence principle” and other duality principles.

While the aim of researchers proposing such principles is to try to “hit upon” fundamental rules of Nature, the fact that principle-based theories start with principles means that we do not have an obvious mechanistic method of deriving what these principles should be. The discovery of new principles is often a matter of familiarity with the data (to allow emergent patterns to be identified) and ... to a certain extent ... the theorist’s personal sense of aesthetics.

Einstein (1922): [6] “The goal of theoretical physics is to create a logical system of concepts based on the fewest possible mutually independent hypotheses, allowing a causal understanding of the entire complex of physical processes.”

A principle, to some extent, represents a declared belief in (or provisional belief in) a rule or set of rules that the universe then might or might not obey. With a principle-based theory, we have the initial principles, the deterministic consequences of not violating those principles, and then usually some leftover issues and implementational details that have to be dealt with by other means.

The purpose of this series of papers [1][5] is to explore a set of three closely related but usually unmentioned principles that appear to run silently through all of Einstein’s work on classical field theory. This is a game with high stakes: if any or all these principles are correct, then we have to consider everything else legitimately built on them to be either true, or near-as-dammit true. If any of these principles is wrong, then they are all wrong, and Einstein’s entire system of classical physics founders.

These three principles are:

A. Symmetry of the Doppler equations (SR as inertial physics),

B. Symmetry of gravitational shifts (GR1916 as gravitational physics), and

C. Symmetry of all physics with respect to time.
2. **The philosophical basis of the current system:**

A. **Doppler symmetry**

A.1 **Assumed absence of any effect of moving matter on light**

In 1905 there were two known ways of reconciling relativity with lightspeed-constancy, a Lorentzian aether, or a dragged aether. ¹

For simplicity, Einstein derived his special theory of relativity by assuming that the behaviour of light transmitted between moving bodies could be exactly defined by assuming that the bodies themselves did not exist (that their presence and behaviour had zero effect on light, always giving flat spacetime), ² and that the behaviour of light in a populated region was therefore exactly defined by what its properties would have been if the region was empty (“Maxwell’s equations of empty space are valid everywhere …”). ³

If the motion of bodies had no effect on light, then the required constant speed of light local to every body could be extrapolated simply throughout the region, giving us global lightspeed constancy (“the universal velocity \(c\)”). ⁴

We then required all inertial observers to agree that the speeds of the same lightsignals were “\(c\)” as measured in all their own different frames,

> **Einstein (1914):** ⁹ “This theory shows that the law of constancy of light propagation in vacuum can be satisfied simultaneously for two observers, in relative motion to each other, such that the same beam of light shows the same velocity to both of them.”

, an apparent impossibility that was resolved by observer-dependent Lorentz redefinitions of distances, times and velocities. ⁵ Different observers disagreed on the distance travelled by the light, but also on the amount of time taken for it to travel:

> **Einstein (1914):** ⁹ “The possibility for such an at first glance paradoxical interpretation can be understood from a more detailed analysis of the physical meaning of spatial and temporal statements. …”

Einstein’s education would have impressed on him the importance of Maxwell’s equations for empty space, the correctness of Lorentzian electrodynamics, and the idea that the possibility of dragged-light models was ruled out by experiment. He already knew what the core mathematics would be assuming global \(c\), because this already appeared under Lorentzian aether theory, which had combined a globally-constant lightspeed with the inability of simple inertial observers to detect a preferred frame. ⁶ Since he already believed that the Lorentz equations were correct, “adopting” a global speed of light from Lorentz’ theory and elevating it to the status of a principle to create the 1905 theory was a natural progression.

---

¹ The other approach being to make \(c\) locally constant for all physical observers, and to have it vary across space as a field in the region between differently-moving masses. ¹² This would be a gravitomagnetic theory ¹³ of local \(c\)-constancy, related to C19th dragged-light aether theories (e.g. Hertz, 1890 ¹⁵).

² Einstein was somewhat coy about acknowledging his 1905 theory’s implicit and unstated third postulate, ¹⁰ that the propagation of light was unaffected by moving matter, perhaps because he did not believe that it would be easily defensible.

³ **Einstein 1914:** ¹¹ “There can be no doubt that this principle is of far-reaching significance; and yet, I cannot believe in its exact validity. It seems to me unbelievable that the course of any process (e.g., that of the propagation of light in a vacuum) could be conceived of as independent of all other events in the world.”

⁵ This approach only works for simple inertial motion in assumed flat spacetime.

⁶ **Einstein (1912):** ⁸ “To fill this gap, I introduced the principle of the [global] constancy of the velocity of light, which I borrowed from H. A. Lorentz’s theory of the stationary luminiferous ether, and which, like the principle of relativity, contains a physical assumption that seemed to be justified only by the relevant experiments.”
A.2 The “bead-in-a-box” scenario

In the “bead-in-a-box” thought-experiment, we shine a laser between two fixed opposing walls of a room, via a glass bead suspended in the beam, and moving along it at velocity \( v \).

a) Newtonian result

Under Newtonian theory, the signal undergoes two transitions, room-to-bead, and bead-to-room, with successive Doppler energy changes of \( E'/E = (c-v)/c \) and \( E'/E = (c-(v))/c \). The total composite shift of the beam received at the far wall is then,

\[
E'/E = \frac{c-v}{c} \times \frac{c+v}{c} = \frac{(c^2-v^2)}{c^2} = 1 - \frac{v^2}{c^2},
\]

a Lorentz-squared redshift after the two frame transitions.

b) SR result, and “shift symmetry”

Special relativity’s required (unstated) third postulate needs the lightsignal to reach the detector at the far side of the room with exactly the same properties, regardless of how the bead moves (or whether or not it exists), so that the total combined shift on the beam, no matter how many beads we use, or how they move, must always be \( E'/E = 1 \).

And this is indeed the outcome with special relativity. With the SR recession Doppler shift characteristic of \( E'/E = \sqrt{(c-v)/(c+v)} \) (Einstein, 1905, §7), we get a total composite shift of

\[
E'/E = \sqrt{\frac{c-v}{c+v}} \times \sqrt{\frac{c+v}{c-v}} = 1
\]

The equation HAS to invert exactly when the polarity of \( v \) is “flipped”, for us to be able to claim that the lightbeam geometry is “flat” as function of velocity, and that the motion of matter has no effect on signal-propagation.

A theory of relativity in flat spacetime has to include this property of shift-symmetry.

A.3 Energy conservation

The implication of this result is that energy conservation also requires shift symmetry, and that relativistic shift symmetry requires special relativity. We could in fact have started with the Newtonian result, pointed out that it describes energy being mysteriously lost for no apparent reason, and derived the “new” relationships of special relativity from the requirement that any credible theory has to conform to the law of conservation of energy.

Since the presence or absence of the moving bead has no effect on the outcome, the Doppler equations do not care whether or not bodies physically exist, and the same results can be calculated using either transitions between bodies, or with purely hypothetical transitions between empty frames. Inertial physics can then be be completely described from just the structural properties of empty (Minkowski) spacetime, the behaviour of matter is completely dictated to by a geometry derived for empty relativistic spacetime, and we have constructed the broad mindset of special relativity.

---

\( i \) ... or a transparent bead made from some crystalline substance, to exploit the Mössbauer method of eliminating standard recoil effects.

\( ii \) The SR velocity addition formula can be derived from the condition that the product of two successive Doppler shifts with velocities \( v_1 \) and \( v_2 \) should be calculable in a single stage, using a composite velocity value \( v \). With the SR shift equations, this \( v \) is not the simple sum of \( v_1 + v_2 \). Since the SR v.a.f. applies regardless of whether is is any actual matter involved, it is considered to be a structural property of Minkowski spacetime.
A.4 Uniqueness of the SR solution regarding shift symmetry

Lorentz aether theory and special relativity together "own" the only relativistic equations for inertial physics that have this special property.

Taking the two most familiar examples of relativistic theories, SR and Newtonian theory, we can note that the Doppler relationships and the theories’ basic definitions both differ between the two systems by a Lorentz factor. We can then define an entire relativistic continuum of hypothetical intermediate relativistic models that each differ from their neighbours by "Lorentzlike" factors, \(^1\) and each will generate a set of the same standard results (such as \(E=mc^2\) and correct muon decay positions), for any model. However, any Lorentzlike deviation away from the predictions of special relativity destroys shift symmetry, as the correction has to be applied equally to approach shifts and recession shifts, which will then no longer cancel.

If we want to combine relativity with global lightspeed constancy, we require spacetime to be "flat", and require the relativistic lightbeam-geometry of a region containing many differently-moving bodies to be identical to the relativistic lightbeam geometry of a region that is empty. Implementing this then requires shift-symmetry.

If we want shift symmetry, the only possible relativistic equations available to use are those of Lorentz aether theory and special relativity.

Special relativity can be defined as (or derived from) the condition of relativity plus shift symmetry.

---

\(^1\) We define a "Lorentzlike" factor as \((1 - \frac{v^2}{c^2})^X\), where the exponent \(X\) is a variable.
B. Gravitational shift symmetry

B.1 GR1916 route-independence of gravitational shifts

Einstein’s *general* theory tends to assume that we can assign the same value to the gravitational differential between two points regardless of the path taken by a signal passing between them (*i.e.*, that the gravitational differential between any pair of points only has one value).

B.2 GR1916 “shift symmetry”

This assumption is again connected to the assumption of energy-conservation. If we aim a beam of light *through* a gravity-well (or through *part* of the gravity-well), we will tend to expect the signal that arrives back at our original height to have the same energy that it started with. This is only possible if the gravitational *blueshift* on a signal passing *downhill* across a gradient with velocity-differential $-v$, and the subsequent *redshift* of the signal then passing *uphill* across the same velocity-differential $+v$, exactly cancel.

If $\text{redshift}(+v) = 1/\text{blueshift}(-v)$

then

$\text{redshift}(+v) \times \text{blueshift}(-v) = 1$

We will refer to this general property as *gravitational shift symmetry*.

B.3 Connection to motion shifts

If the gravitational differential between two points linked by a geodesic causes freefalling body to change velocity by $v$, and the increase or decrease in energy of a system making that journey is the same regardless of whether the system’s massenergy makes the journey as matter or as some other form of energy, then we can *define* the gravitational differential in terms of the equivalent resulting velocity-change $v$, and calculate the gravitational shift directly from the expected final Doppler shift, red or blue, for a mass that has acquired this new velocity of $\pm v$ after free-falling across the differential.

The condition of exact energy-conservation (in the way that the problem is stated here) then requires us to use a Doppler shift relationship in inertial physics where the redshift for a body receding at $v$ is the exact inverse of the blueshift for a body approaching at $v$.

In other words, shift symmetry in *gravitational* physics requires shift symmetry in *inertial* physics.

B.4 GR1916 and the SR legacy

This condition is conveniently met by special relativity, whose Doppler relationship is the symmetrical $E'/E = \sqrt{(c-v)/(c+v)}$ (Einstein, 1905, [*17*]) A moment’s consideration shows that if we invert the sign of $v$, this inverts the right-hand side of the equation, and therefore also the ratio $E'/E$.

Over a gravitational round-trip, we then obtain

$$\frac{E'}{E} = \sqrt{\frac{c-v}{c+v}} \times \sqrt{\frac{c-(v)}{c+(v)}} = 1$$
B.5 The Schwarzschild solution

Since so much of general relativity deals with abstract quantities and properties, and the theory’s physical predictions are so often only given as approximations (often Newtonian approximations), it can be difficult to find a definitive statement of the theory’s actual exact prediction for the spectral shift on a signal moving between two specified locations. Karl Schwarzschild’s 1916 solution to Einstein’s system is considered exact, and Robert M. Wald’s “General Relativity” [191] comes to our rescue by translating the Schwarzschild solution into a gravitational shift prediction.

Wald’s equation 6.3.5 gives the Schwarzschild solution’s prediction of gravitational shifts as

\[
\frac{\omega_1}{\omega_2} = \left(\frac{1-2M/r_2}{1-2M/r_1}\right)^{1/2} \quad \ldots \quad 6.3.5
\]

, where \(\omega_1/\omega_2\) is the ratio of frequencies, and \(r_1\) and \(r_2\) are the two nominal radii that the signal passes between. Without studying the rest of the equation, it is immediately obvious that swapping the radii \(r_1\) and \(r_2\) inverts the frequency ratio: sending a signal from \(r_1\) to \(r_2\) and then back from \(r_2\) to \(r_1\) results in a perfect cancellation of the combined redshifts and blueshifts, so that \(E'/E\), over the round trip, equals “one”. 1

In other words, the Schwarzschild solution is shift-symmetrical.

For an observer at null infinity, a signal generated at \(r=2M\) has a calculated received frequency of zero, and for a signal sent the other way, the calculated received frequency is infinite. ii iii iv v vi vii viii

---

i See also Equation 2.19 in Theory and Experiment in Gravitational Physics (page 33). [20]

ii MTW (1974) [21] “§7.2. Gravitational Redshift Derived from Energy Conservation” invokes a more a more limited form of energy-conservation: that the energy we get out of a system shouldn’t be any more than we put in, (i.e. "No infinite energy machines").

iii Einstein’s famous 1911 paper on gravity-shifts [22] does not give the GR1916 predictions (the full general theory having not yet been completed), or even the predictions required for compatibility with special relativity. Instead it gives the corresponding calculations for the necessary change on energy of falling light, if Newtonian theory is correct rather than SR. As Einstein qualifies:

Einstein (1911): [22] “The relations here deduced, even if the theoretical basis is sound, are valid only to a first approximation. ... To avoid unnecessary complications, let us for the present disregard the theory of relativity, and regard both systems from the customary point of view of kinematics, and the movements occurring in them from ordinary mechanics. ...”

iv It is often difficult to get a straight answer from GR sources as to the exact GR prediction for gravitational shifts. The "Newtonian approximation" \(gh\) is often used instead (treated as the approximate GR prediction "for testing purposes"), but for exact theoretical work, this would give a Newtonian "dark star" rather than a Wheeler black hole.

v Some GR physics folk will insist on arguing that the correct GR prediction is just \(1+gh\), despite Einstein’s 1911 qualification that this was a Newtonian, "pre-SR" calculation. The "Newtonian" gravitational shift on falling light-corpuscles was published by John Michell back in 1784. [23] If the GR prediction really was the same as the SR prediction, it would have been difficult for Einstein to pass off the GR redshift prediction as being one of the fundamental tests of GR, and it would have been difficult for theorists to characterise the GR shift as being a new and non-Newtonian effect.

Assuming that the relativity community is competent and honest, we must assume that the GR1916 prediction is in some way different from the Newtonian, and the simplest way for this to be true is if the exact GR1916 relationships are as stated here (and as given by Wald [19]).

vi Shutz’ second edition (2009) [24] gives a GR frequency shift of:

\[
\frac{\text{[nu]}(\text{freely falling})}{\text{[nu]}(\text{apparatus at top})} = 1 + gh + 0(\nu') \quad \ldots \quad (5.2)
\]

Since \(gh\) is velocity, the \(1+gh\) is equivalent to the Newtonian Doppler shift of \(E/E'-(c\cdot\nu)/c\), which over a round trip would give an energy loss of \(1-\nu'/c^2\). However, the \(0(\nu')\) term is introduced earlier as a Lorentz factor ("transverse component"), presumably used to blueshift the Newtonian prediction to bring it into line with special relativity, cancelling what would otherwise be the Newtonian Lorentz-squared redshift.

vii The Newtonian \(1+gh\) calculation does NOT cancel over a round-trip, as the "\(gh\)" is not simply added and then subtracted again: rather, the two frequency ratios for the two journeys must be multiplied together, giving \((1+gh) \times (1-gh) = 1 - (gh)^2\) .

viii The gravitational equations also obviously have to be the SR set in order for the resulting system to be time-symmetrical.
C. Time-symmetry

C.1 Eddington on time

Arthur Eddington was one of Einstein’s most important supporters, and his 1927 Gifford lecture[^25] became one of the most influential Twentieth-Century explorations of the concept of time.

Eddington (1929):[^25] “Time’s Arrow. ... In the four-dimensional world considered in the last chapter the events past and future lie spread out before us as in a map. The events are there in their proper spatial and temporal relation; but there is no indication that they undergo what has been described as “the formality of taking place”, and the question of their doing or undoing does not arise. We see in the map the path from past to future or from future to past; but there is no signboard to indicate that it is a one-way street. Something must be added to the geometrical conceptions comprised in Minkowski’s world before it becomes a complete picture of the world as we know it. We may appeal to consciousness to suffuse the whole — to turn existence into happening, being into becoming. But first let us note that the picture as it stands is entirely adequate to represent those primary laws of Nature which, as we have seen, are indifferent to a direction of time. ...”

Eddington continues,

“... Objection has sometimes been felt to the relativity theory because its four-dimensional picture of the world seems to overlook the directed character of time. The objection is scarcely logical, for the theory is in this respect no better and no worse than its predecessors. The classical physicist has been using without misgiving a system of laws which do not recognise a directed time; he is shocked that the new picture should expose this so glaringly. ”

Eddington acknowledges that special relativity has no arrow of time, but insists that this should not be taken as a criticism of SR, as all previous classical systems had been time-symmetrical, too.[^1] ... according to Eddington, special relativity had merely made the existing, implicit time-symmetries more obvious.

---

[^1]: ... which, we’ll find in Paper C [4] is not correct.
C.2 Uniqueness of the SR solution regarding time symmetry

Special relativity’s shift symmetry and GR1916’s gravitational shift symmetry also translate into time symmetry – reversing the arrow of time inverts the polarity of v (“flipping” the Doppler equation), and also reverses the ratio E’/E, so that the relationship is the same in both forward and reversed time.

Since the slightest Lorentzlike deviation from the SR relationships breaks time-symmetry, if we require relativistic physics to be time-symmetrical, the SR equations and Einstein’s structure are our only choice. i ii iii

---

i A Lorentzlike factor takes the form \((1 - \frac{v^2}{c^2})^X\). We can define a continuum of candidate relativistic equations (incorporating both SR and C19th Newtonian optics, “NO”), where every solution differs from its neighbours in its predictions and definitions by a Lorentzlike factor. If we pick any point on this continuum as our initial reference (say, SR or NO), the location and properties of any other solution can be defined by a single parameter, the exponent “X” of the theory’s Lorentzlike deviation from the chosen reference-theory.

ii If we chose to use SR as our reference, we can see that any Lorentzlike deviation from SR breaks SR’s time-symmetry. SR is the only symmetrical solution to relativity theory.

iii … we can therefore “prove” special relativity from just the assumption of the relativity principle, combined with Eddington’s statement that all classical theories are time-symmetrical.
3. **Apparent impregnability of the SR-GR1916 system**

At this point, we have recreated the prevailing mindset and design logic that applied to Einstein’s work towards a general theory, and its SR foundation, and most follow-on work, from around 1905 through the rest of the Twentieth Century. It seemed quite obvious that gravitational theory had to obey traditional energy-conservation, and needed to be time-symmetrical, and we could prove without a doubt that these things could only be true if the equations of inertial physics were those of special relativity. Special relativity therefore seemed to arise naturally from Einstein’s general theory, even if one did not explicitly write its definitions into the theory (as Einstein had), or try to invoke a geometrical reduction to SR. [26]

It seemed absolutely impossible and inconceivable that the system could be wrong in any way, short of any new behaviours that might appear at tiny scales due to quantum theory.

4. **Conclusions**

Einstein’s belief system and sense of aesthetics in theoretical physics can be likened to a platform standing on three pillars, each of which is cross-braced by the other two. Additionally, each pillar, in isolation, can be used to regenerate the other two: if we believe in symmetrical inertial physics, we get symmetrical gravitational physics, and also time-reversibility; if we believe in gravitational shift-symmetry and the Schwarzschild solution, we can derive special relativity’s equations from the motion shift of a falling body, and again, get time-reversibility; if we assume time-reversibility, the equations for inertial and gravitational physics must be exactly as Einstein described them.

If any one of the three pillars is correct, they must all be correct. Conversely, for any of them to be wrong, they must all be wrong.

Having apparently shown that Einstein’s framework cannot possibly be wrong, it is the job of the rest of this sequence of papers to explain how and why all three sets of arguments are problematic to the point of unusability, that geometrical frameworks suitable for physics have to instead be asymmetrical, and that Einstein’s mathematically “neat and tidy” system is not, after all, a suitable basis for physical law. i

---

i Richard Feynman argued that it was the responsibility of the scientist to go out of their way to identify any possible failings in their own theory, and to be scrupulously fair to their opposition. Many researchers do not do this: Einstein, in particular seems to have considered scientific debate as being akin to gladiatorial combat or a court of law, where it was the responsibility of the advocate to paint their opponents in the worst possible light, and if any linguistic or logical sleight-of-hand made the case look unfairly-favourable for their own side, then it was their opponents’ job to point it out.

For this sequence of papers, we are following the “Feynman rules”, and are therefore starting by deliberately stating the case for Einstein’s system in what seems to be the clearest, most efficient, most easily understandable, and most compelling way possible. The reasoning in this first paper is arguably more convincing than that found in mainstream sources promoting Einstein’s system, including some of Einstein’s own.

Since the destruction of the Einstein “symmetry” worldview (in Papers A-C onwards), is logical-geometrical and apparently unavoidable, we have no reason to be anything but scrupulously fair to the system that we are about to dismantle. The arguments to be presented are sufficiently strong that they do not have to hide behind ambiguity or misdirection.
References


23. John Michell, "On the means of discovering the distance, magnitude, &c. of the fixed stars, in consequence of the diminution of the velocity of their light ...", *Philosophical Transactions of the Royal Society of London* **vol. 74** (1784), pages 35-57. [https://doi.org/10.1098/rstl.1784.0008](https://doi.org/10.1098/rstl.1784.0008)


---