Rev. Dec. 13,. Jan 17, 2004; December 1; November 29, 2003

Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies that Led him to it.

John D. Norton[1]
Department of History and Philosophy of Science
University of Pittsburgh
Pittsburgh PA 15260
jdnorton@pitt.edu

Prepared for *Cambridge Companion to Einstein*, M. Janssen and C. Lehner, eds.,
Cambridge University Press.

Typesetter: the figures embedded here are lower resolution tif images. The original figures were drawn using a vector graphics program (Canvas 9) and should be printed from vector graphic files, e.g. epsf.

1. Introduction

Modern readers turning to Einstein's famous 1905 paper on special relativity may not find what they expect. Its title, "On the electrodynamics of moving bodies," gives no inkling that it will develop an account of space and time that will topple Newton's system. Even its first paragraph just calls to mind an elementary experimental result due to Faraday concerning the interaction of a magnet and conductor. Only then does Einstein get down to the business of space and time and lay out a new theory in which rapidly moving rods shrink and clocks slow and the speed of light becomes an impassable barrier. This special theory of relativity has a central place in modern physics. As the first of the modern theories, it provides the foundation for particle physics and for Einstein's general theory of relativity; and it is the last point of agreement between them. It has also received considerable attention outside physics. It is the first port of call for philosophers and other thinkers, seeking to understand what Einstein did and why it changed everything. It is often also their last port. The theory is arresting enough to demand serious reflection and, unlike quantum theory and general relativity, its essential content can be grasped fully by someone merely with a command of simple algebra. It contains Einstein's analysis of simultaneity, probably the most celebrated conceptual analysis of the century.

Many have tried to emulate Einstein and do in their fields just what Einstein did for simultaneity, space and time. For these reasons, many have sought to understand how Einstein worked his magic and came to special relativity. These efforts were long misled by an exaggeration of the importance of one experiment, the Michelson-Morley experiment, even though Einstein later had trouble recalling if he even knew of the experiment prior to his 1905 paper.[2] This one experiment, in isolation, has little force. Its null result happened to be fully compatible with Newton's own emission theory of light. Located in the context of late 19th century electrodynamics when ether-based, wave theories of light predominated, however, it presented a serious problem that exercised the greatest theoretician of the day.

Another oversimplification pays too much attention to the one part of Einstein's paper that especially fascinates us now: his ingenious use of light signals and clocks to mount his conceptual analysis of simultaneity. This approach gives far too much importance to notions that entered briefly only at the end of years of investigation. It leaves us with the curious idea that special relativity arrived because Einstein took the trouble to think hard enough about what it means to be simultaneous. Are we to believe that the generations who missed Einstein's discovery were simply guilty of an oversight of analysis?[3] Without the curious behavior of light, as gleaned by Einstein from 19th century electrodynamics, no responsible analysis of clocks and light signals would give anything other than Newtonian results.

Why did special relativity emerge when it did? The answer is already given in Einstein's 1905 paper. It is the fruit of 19th century electrodynamics. It is as much the theory that perfects 19th century electrodynamics as it is the first theory of modern physics. [4] Until this electrodynamics emerged, special relativity could not arise; once it had emerged, special relativity could not be stopped. Its basic equations and notions were already emerging in the writings of H. A. Lorentz and Henri Poincaré on electrodynamics. The reason is not hard to understand. The observational consequences of special relativity differ significantly from Newtonian theory only in the realm of speeds close to that of light. Newton's theory was adapted to the fall of apples and the slow orbits of planets. It knew nothing of the realm of high speeds. Nineteenth century electrodynamics was also a theory of light and the first to probe extremely fast motions. The unexpected differences between processes at high speeds and those at ordinary speeds were fully captured by the electrodynamics. But their simple form was obscured by elaborate electrodynamical ornamentations. Einstein's achievement was to strip them of these ornamentations and to see that the odd behavior of rapidly moving electrodynamical systems was not a peculiarity of electricity and magnetism, but imposed by the nature of space and time on all rapidly moving systems.

This chapter will present a simple statement of the essential content of Einstein's special theory of relativity, including the inertia of energy, $E=mc^2$. It will seek to explain how Einstein extracted the theory from electrodynamics, indicating the subsidiary roles played by both experiment and Einstein's conceptual analysis of simultaneity.

All efforts to recount Einstein's path face one profound obstacle, the near complete lack of primary source materials. This stands in strong contrast to the case of general relativity, where we can call on a seven year record of publication, private calculations and an extensive correspondence, all prior to the completion of the theory. (See General Relativity, this volume.) For special relativity, we have a few fleeting remarks in Einstein's correspondence prior to the 1905 paper and brief, fragmented recollections in later correspondence and autobiographical statements. The result has been an unstable literature, pulled in two directions. The paucity of sources encourages accounts that are so lean as to be uninformative. Yet our preoccupation with the episode engenders fanciful speculation that survives only because of the lack of source materials to refute it. My goal will be an account that uses the minimum of responsible conjecture to map paths between the milestones supplied by the primary source materials.

2. Basic Notions

2.1 Einstein's postulates

Einstein's special theory of relativity is based on two postulates, stated by Einstein in the opening section of his 1905 paper. The first is the *principle of relativity*. It just asserts that the laws of physics hold equally in every inertial frame of reference.[5] That means that any process that can occur in one frame of reference according to these laws can also occur in any other. This gives the important outcome that no experiment in one inertial frame of reference can distinguish it intrinsically from any other. For that same experiment could have been carried out in any other inertial frame with the same outcome. The best such an experiment can reveal is motion with respect to some other frame; but it cannot license the assertion that one is absolutely at rest and the other is in true motion.

While not present by name, the principle of relativity has always been an essential part of Newtonian physics. According to Copernican cosmology, the earth spins on its axis and orbits the sun. Somehow Newtonian physics must answer the ancient objection that such motions should be revealed in ordinary experience if they are real. Yet, absent astronomical observations, there is no evidence of this motion. All processes on earth proceed just as if the earth were at rest. That lack of evidence, the Newtonian answers, is just what is expected. The earth's motions are inertial to very good approximation; the curvature of the trajectory of a spot on the earth's surface is small, requiring 12 hours to reverse its direction. So, by the conformity of Newtonian mechanics to the principle of relativity, we know that all mechanical processes on the moving earth will proceed just as if the earth were at rest. The principle of relativity is a commonplace of modern life as well. All processes within an airplane cabin, cruising rapidly but inertially, proceed exactly as they would at the hangar. We do not need to adjust our technique in pouring coffee for the speed of the airplane. The coffee is not left behind by the plane's motion when it is poured from the pot.

Einstein's second postulate, the *light postulate*, asserts that "light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body." Einstein gave no justification for this postulate in the introduction to his paper. Its strongest justification came from Maxwell's electrodynamics. That theory had identified light with waves propagating in an electromagnetic field and concluded that just one speed was possible for them in empty space, $c = 300{,}000$ km/sec, no matter what the motion of the emitter.

2.2 Relativity of simultaneity

Einstein pointed out immediately that the two postulates were "apparently irreconcilable." His point was obvious. If one inertially moving observer measures *c* for the speed of some light beam, what must be measured by another inertially moving observer who chases after the light beam at high speed—say 50% of *c* or even 99% of *c*? That second observer must surely measure the light beam slowed. But if the light postulate respects the principle of relativity, then the light postulate must also hold for this second, inertially moving observer, who must still measure the same speed, *c* for the light beam.

How could these conflicting considerations be reconciled? Einstein's solution to this puzzle became the central conceptual innovation of special relativity. Einstein urged that we only think the two postulates are incompatible because of a false assumption we make tacitly about the simultaneity of events separated in space. If one inertially moving observer judges two events, separated in space, to be simultaneous, then we routinely assume that any other observer would agree. That is the false assumption. According to Einstein's result of the *relativity of simultaneity*, observers in relative motion do not agree on the simultaneity of events spatially separated in the direction of their relative motion.

To demonstrate this result, Einstein imagined two places A and B, each equipped with identically constructed clocks, and a simple protocol to synchronize them using light signals. In simplified form, an observer located at the midpoint of the platform holding A and B waits for light signals emitted with each clock tick. The observer would judge the clocks properly synchronized if the signals for the same tick number arrive at the observer at the same time, for the signals propagate at the same speed *c* in both directions. The check of synchrony is shown in Figure 1, where the platform at successive times is displayed as we proceed up the page.

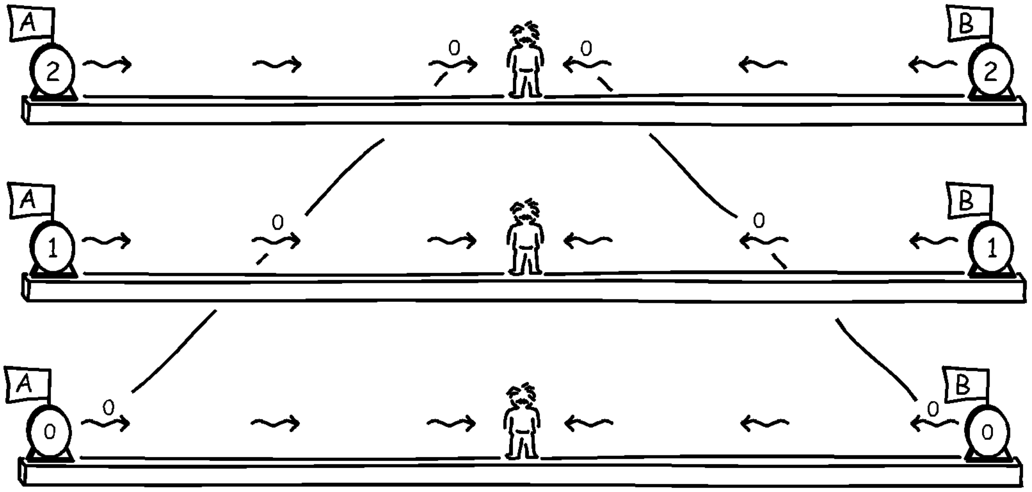

Figure 1 Checking the synchrony of two clocks

Now imagine how this check of synchrony would appear to another observer who is moving inertially to the left and therefore sees the platform move to the right. To this observer, the fact that the two zero-tick signals arrive at the same time is proof that the two clocks are *not* properly synchronized. For the moving observer would judge the platform observer to be rushing away from clock A's signal and rushing towards that of clock B. So signals emitted by clock A must travel further to reach the platform observer O than signals emitted by clock B. The moving observer would judge the zero-tick of clock A to occur before the zero tick of clock B; and so on for all other ticks. The light postulate is essential for this last step, which depends upon the moving observer *also* judging light signals in both directions to propagate at *c*; without this postulate, the relativity of simultaneity cannot be derived.

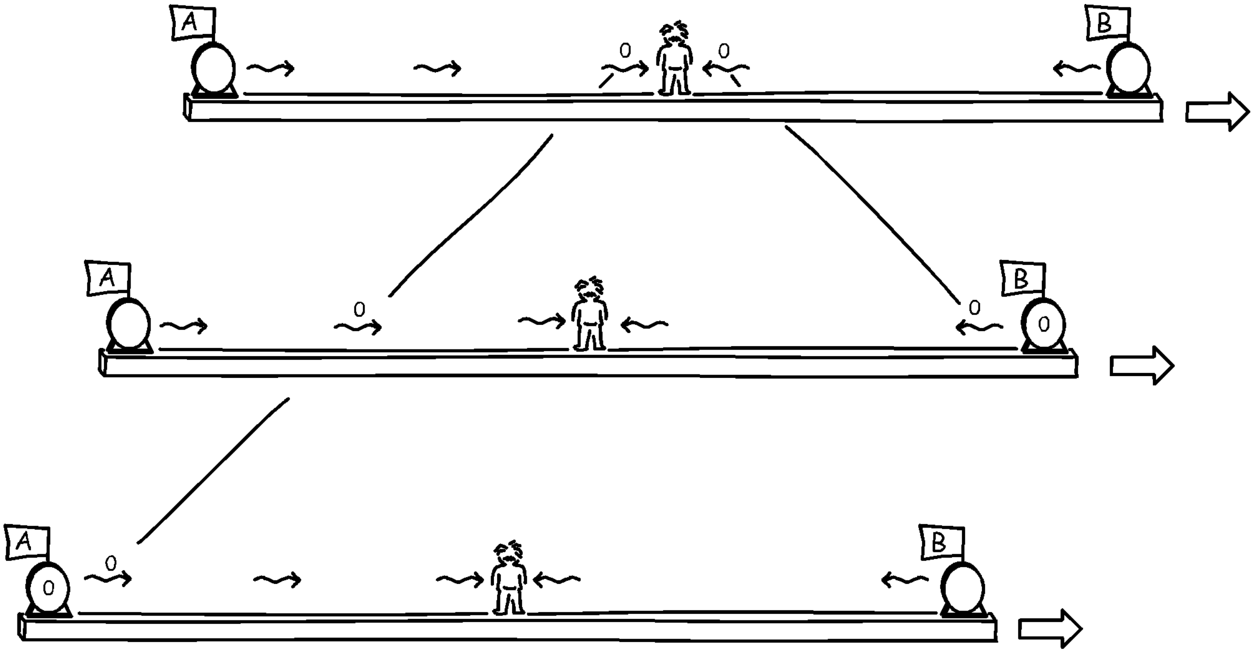

Fig 2. Check of clock synchrony as seen by a moving observer

Since observers can use clocks to judge which events are simultaneous, it now follows that they disagree on which pairs of events are simultaneous. The platform observer would judge the events of the zero tick on each of clocks A and B to be simultaneous. The moving observer would judge the zero tick on clock A to have happened earlier.

This simple thought experiment allows us to see immediately how it is possible for Einstein's two postulates to be compatible. We saw that the constancy of the speed of light led to the relativity of simultaneity. We merely need to run the inference in reverse. Let us make the *physical assumption* that space and time are such that clocks are in true synchrony when set by the above procedure. Then, using properly synchronized clocks in our frame of reference, whichever it may be, we will always judge the speed of light to be *c*. Suppose we chase after a light signal, no matter how rapidly. Since we will have changed frames of reference, we will need to resynchronize our clocks. Once we have done that, we will once again measure a speed *c* for the light signal.

2.3 Kinematics of special relativity

Much of the kinematics of special relativity can be read from the relativity of simultaneity. One effect can be seen in the figures above. Figure 1 shows that the platform observer will judge there to be as many light signals moving from left to right over the platform as from right to left. A direct expression of the relativity of simultaneity is that the moving observer will judge there to be more signals traversing from A to B, laboriously seeking to catch the fleeing end of the platform; while there will be fewer traversing from B to A, since they approach an end that moves to meet them.

To see another effect, imagine that the horizontal platform moves vertically and that it passes horizontal lines, aligning momentarily with each as it passes, as shown in Figure 3.

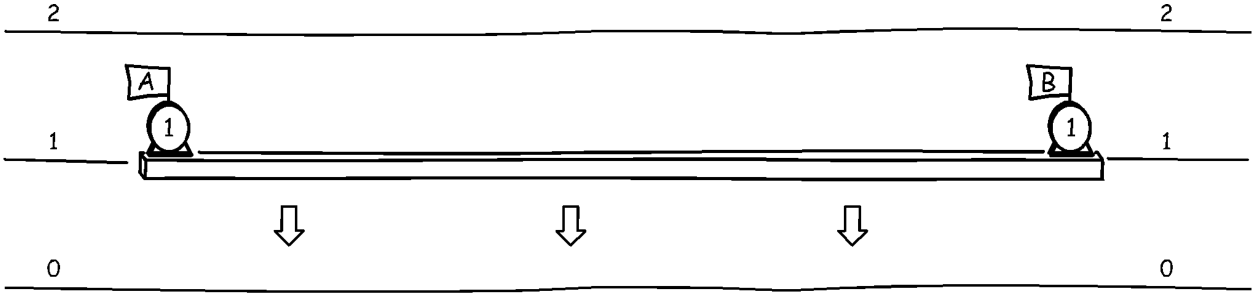

Figure 3. Vertical motion

That alignment depends on judgments of simultaneity: that the event "A passes line 1" is simultaneous with the event "B passes line 1," for example. Another observer who also judges the platform to move to the right would *not* judge these two events to be simultaneous. That observer would judge the A event to occur before the B event. The outcome, as shown in Figure 4, is that the horizontal motion would tilt the platform so that it is no longer horizontal. That rotation is a direct expression of the relativity of simultaneity. A manifestation of this rotation arises in stellar aberration, discussed below in Section 4.5.

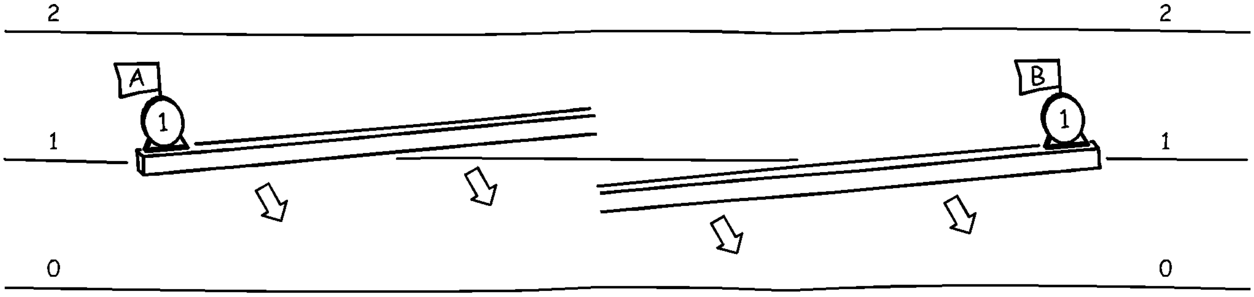

Figure 4. Vertical motion seen by a horizontally moving observer

The more familiar kinematical effects of special relativity also follow from the relativity of simultaneity simply because the measurement of any property of a moving process requires a judgment of simultaneity. For example, we may measure the length of a rapidly moving car by placing two marks *simultaneously* on the roadway as the car passes, one aligned with the front and one with the rear. We then measure the distance between the marks to determine the length of the car. Or we may judge how fast the car's dashboard clock is running by comparing its readings with those of synchronized clocks we have laid out along the roadway. A straightforward analysis would tell us that the rapidly moving car has shrunk and its clock slowed. The car driver would not agree with these measurements since they depend upon our judgment of the simultaneity of the placing of the marks and synchrony of the clocks. Indeed the car driver, carrying out an analogous measurement on us would judge that our rods have shrunk and our clocks have slowed—and by the same factors, just as the principle of relativity demands.

That we each judge the other's rods shrunk and clocks slowed is typical of relativistic effects. At first they seem paradoxical until we analyze them in terms of the relativity of simultaneity. Most complaints that relativity theory is paradoxical derive from a failure to accept the relativity of simultaneity.

The full complement of these kinematical effects is summarized in the equations of the *Lorentz transformation*. They describe what transpires when we view a system from two different inertial frames of reference; or, equivalently, what happens to one system when it is set into inertial motion. The body shrinks in length in the direction of motion; all its temporal processes slow; and the internal synchrony of its parts is dislocated according to the relativity of simultaneity. All these processes approach pathological limits as speeds approach c, which functions as an impassable barrier. The Lorentz transformation was not limited to spaces and times. Just as spaces and times transform in unexpected ways, Einstein's analysis of electrodynamical problems depended on an unexpected transformation for electric and magnetic fields. As we change inertial frames, a pure electric field or pure magnetic field may transform into a mixture of both.

The classical analog of the Lorentz transformation was later called the *Galilean transformation*. According to it, moving bodies behave just as you would formerly have expected: motion does not alter lengths, temporal processes or internal synchrony and there is no upper limit to speeds.

A mathematically perspicuous representation of Einstein's kinematics was given by Hermann Minkowski in 1907 in terms of the geometry of a four-dimensional spacetime. It lies outside the scope of this chapter.

3. Lorentz's Theorem of Corresponding States

3.1 Failing to see the ether wind[6]

While Newton's physics had conformed to the principle of relativity, the revival of the wave theory of light in the early 19th century promised a change. Light was now pictured as a wave propagating in a medium, the luminiferous ("light bearing") ether, which functioned as a carrier for light waves, much as the air does for sound waves. It seemed entirely reasonable to expect that this ether would provide the state of rest prohibited by the principle of relativity. As the earth moves through space, a current of ether must surely blow past. A series of optical experiments were devised to detect the effects of this ether wind. The curious result in experiment after experiment was that no such result could be found. All "first order" experiments, that is, ones that required the least sensitivity of the apparatus, yielded a null result.[7] This failure could be explained by a simple result, the Fresnel ether drag. The speed of light in an optically dense medium (like glass) with refractive index n is c/n. What would the speed of the light be if that medium moves with some speed v in the same direction? Will that speed be fully added to that of light? Fresnel proposed that only a portion would be added, precisely $v(1–1/n^2)$, imagining that the ether is partially dragged by the medium. It has to be just that factor. It turns out that if the ether is dragged by just that amount, then no first order experiment can reveal the ether wind.

By the middle of the 19th century, the problem was enlarged by Maxwell's discovery that light was actually a wave propagating in the electromagnetic field. Maxwell's theory was also based on an ether that carried the electric and magnetic fields of his theory and it too supplied a state of rest prohibited by the principle of relativity. The problem of explaining why no ether wind was detectable became part of a larger problem in electrodynamics. It became more acute when the Michelson-Morley experiment of 1887, the first second order experiment, detected no ether wind. By 1903, Trouton and Noble had carried out a fully electrodynamic second order experiment, again with a null result. (See Janssen, 1995, Ch. 1.)

3.2 A challenging problem in electrodynamics

The task of accommodating electrodynamics to these null results was undertaken by the great Dutch physicist, Henrik A. Lorentz. In a series of papers in the 1890s and early 1900s, he was able to show that Maxwell's electrodynamics should not be expected to yield any positive result in these experiments. The computational task he faced was formidable. To arrive at his result, he needed a systematic comprehension of moving systems in electrodynamics. Motion immensely complicates electrodynamics. Take, for example, the basic entity of his electrodynamics, the electron, which he modeled as a sphere of electric charge surrounded by an electric field **E**. As long as it is at rest in the ether, it could be analyzed merely by looking at the electrostatic forces between each of the parts of the electron. But once the electron is set in motion through the ether, each part becomes a moving charge; and a moving charge is an electric current; and an electric current generates a magnetic field **H**; and that magnetic field acts on moving charges. See Figure 5. A thorough analysis is messy and eventually shows that the electron must be contracted slightly in its direction of motion.

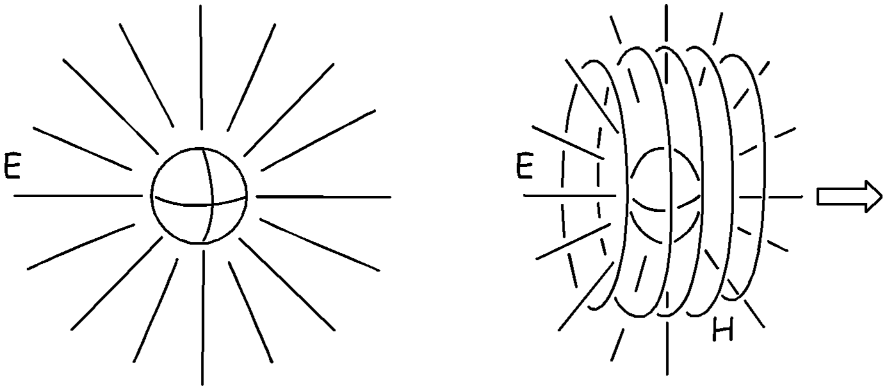

Figure 5. Lorentz's electron at rest and in motion

The problem of computing the behavior of moving systems had been immeasurably easier in Newtonian physics since it conformed to the principle of relativity. The principle could be used to convert hard problems in moving systems into easy problems in systems at rest. Suppose, for example, that that we want to know if a rapidly moving asteroid can gravitationally capture a satellite. What initial speed should we give the satellite so that capture is possible?

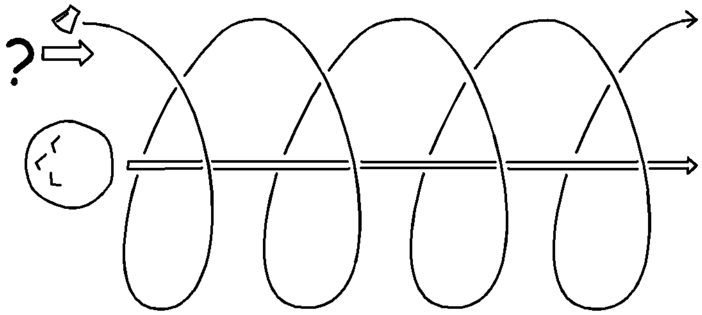

Figure 6. A hard problem in Newtonian physics

The problem is solved by first solving a much easier problem: if the asteroid were at rest, is such a capture possible? Obviously, yes. What initial speed is needed? Computing it is the easiest problem in celestial mechanics.

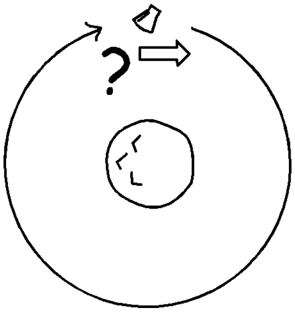

Figure 7. An easy problem in Newtonian physics.

But once we have solved the easy problem, we have also solved the hard problem, for the principle of relativity tells us that we recover a full description of a moving asteroid with its satellite by merely taking the easy case of the asteroid at rest and setting it into uniform motion by means of a Galilean transformation.

3.3 The theorem[8]

What Lorentz needed urgently was some computational device like the principle of relativity so he could find solutions of Maxwell's equations easily for moving systems. But Maxwell's electrodynamics does not conform to the principle of relativity. Its equations hold only in a frame of reference at rest in the ether. Lorentz's ingenious discovery was a theorem in Maxwell's electrodynamics that mimicked the principle of relativity sufficiently for his purposes. The principle of relativity says that one can generate new systems compatible with the laws of nature by taking one solution and constructing *identical* uniformly moving copies. Lorentz saw that essentially the same thing could be done with Maxwell's electrodynamics. One could start with a solution of Maxwell's equations and produce oddly *distorted* moving copies of them. If one used just the right distortions, one would be assured that the new systems, the "corresponding state" of the old system, would also solve Maxwell's equations.

The rules Lorentz specified should not be a surprise. They are just the Lorentz transformation described above in Section 2. But Lorentz did not give them Einstein's interpretation. They were merely artifices whose quite odd form was fixed by Maxwell's equations and justified solely by the fact that they enabled construction of new solutions from old. The largest (first order) effect was a dislocation of the internal synchrony of the parts of the system that we now know as the relativity of simultaneity. For Lorentz, the rule was simply the assembly of a new system from the parts of the old, sampled at different times. The sampling rule was governed by his notion of "local time"—a sampling time that varied with the spatial location (hence "local"). Other first order effects included odd transformations of fields: a pure electric field, such as the one surrounding an electron at rest, would become a mixture of electric and magnetic fields, just as shown in Figure 5. This first order transformation was developed in Lorentz's (1895) *Versuch*. Higher order effects soon followed and were codified in Lorentz (1904) They included the slowing of all temporal processes and the contraction of lengths in the direction of motion. (Einstein did not know of this later paper when he wrote his own on special relativity.)

With these rules and his theorem, Lorentz was able to compare systems moving and at rest in the ether and show that no existing experiment could decide which was at rest and which was moving. His device of local time was adequate for all first order experiments, including the recovery of the Fresnel drag coefficient (though not the interpretation of a dragged ether). The higher order contraction was sufficient for the Michelson-Morley experiment.

We can see just how Lorentz used these rules to describe electrons in motion. He solved the easy problem of electrons at rest and used the transformation to form its corresponding state, a contracted moving electron surrounded by a magnetic field. This example reveals an important complication. The electron at rest in Figure 5 cannot be governed solely by electromagnetic forces. Since like charges repel, another otherwise unknown, non-electromagnetic force must be present in order to hold all the parts of the electron together and prevent it blowing itself apart. How might this force transform? Lorentz made the natural supposition that it would transform just like electric and magnetic forces do under his Lorentz transformation. Only then could the contracted, moving electron of Figure 5 be recovered. This was a weak point of Lorentz's account for he was required to make presumptions about forces whose nature was quite unknown to him. The resulting contraction also happens to be the same length contraction used to explain the Michelson-Morley experiment, where it is sometimes called the Lorentz-Fitzgerald contraction. The awkwardness surrounding its introduction has led to suggestions that Lorentz's account is *ad hoc*. A better assessment is given by Janssen (2002, 2002a), who urges that the superiority of Einstein's treatment lies in its giving a single explanation for what is otherwise an odd coincidence. Einstein shows us that forces of all types must transform alike because they inhabit the same space and time.

4. Einstein's Path to Special Relativity[9]

4.1 The magnet and conductor thought experiment

The decisive moment in Einstein's path to special relativity came when he reflected on the interaction of a magnet and conductor in Maxwell's electrodynamics. The outcome was of such enduring importance that, years later when he wrote his 1905 paper on special relativity, this was the elementary consideration to which he gave pride of place in the paper's first paragraph.[10] As far as Maxwell's theory is concerned, the case of a magnet at rest in the ether is very different from that of one that moves. As shown in Figure 8 the magnet at rest is surrounded just by a static magnetic field **H**.

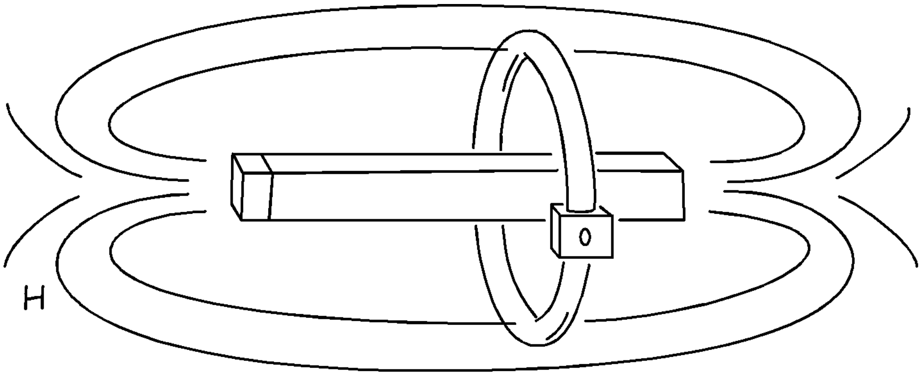

Figure 8. Magnet and conductor at rest in the ether

The moving magnet, however, is surrounded by both a magnetic field **H** and an electric field **E**. The latter arises from the complicated interactions between electric and magnetic fields in Maxwell's electrodynamics. At a point in space as the magnet moves past, the magnetic field will wax and wane. A time varying magnetic field induces an electric field, a new entity not present in the first case.

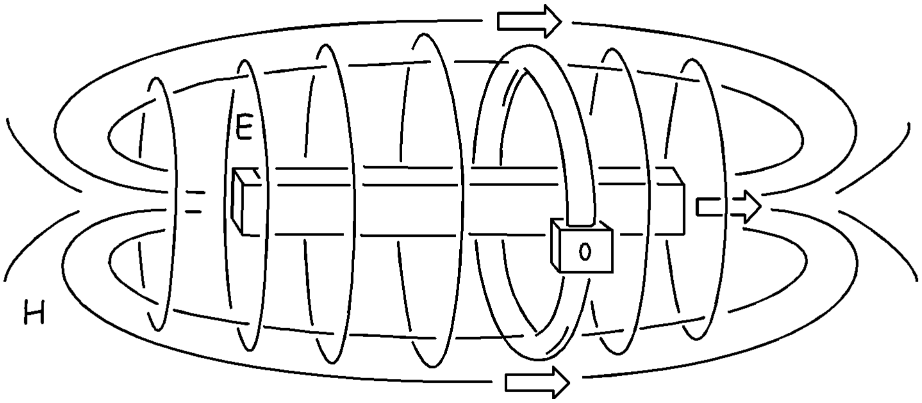

Figure 9. Magnet and conductor moving in the ether

Since the theory holds the two cases to be so distinct, one would expect that a simple measurement would distinguish them. The most straightforward would be to encircle the magnet with a conductor; that is, a wire with free charges in it that would be set in motion by the electric field to generate a measurable electric current. The conductor surrounding the magnet at rest would show no current; the conductor moving with the moving magnet would show a current and reveal its absolute motion. Or so one would expect. However another electrodynamical interaction intervenes. Since the charges of the moving conductor are themselves moved through the magnetic field, that field also exerts a force on them and produces a current. The two currents—one due to the induced electric field, the other due to the motion of the charges in the magnetic field—are in opposite directions and turn out to cancel exactly. In both cases, there is no measurable current. Once again we have an experiment aimed at detecting motion in the ether, this time using a simple detector made from a magnet and a wire. And again we find a null result.

Einstein (1920) later recalled how disturbed he was by the tension between the theoretical account and experimental outcome:

> The idea, however, that these were two, in principle different cases was unbearable for me. The difference between the two, I was convinced, could only be a difference in choice of viewpoint and not a real difference. Judged from the [moving] magnet, there was certainly *no* electric field present. Judged from the [ether], there certainly was one present. Thus the existence of the electric field was a relative one, according to the state of motion of the coordinate system used, and only the electric and magnetic field *together* could be ascribed a kind of objective reality, apart from the state of motion of the observer or the coordinate system. The phenomenon of magneto-electric induction compelled me to postulate the (special) principle of relativity.

The principle of relativity, which prevailed among the observables, had to be extended to the full theory. This thought experiment gave Einstein the means to do it. The existence of the induced electric field was no longer the immutable mark of a magnet truly in motion; it was now merely an artifact of motion relative to the observer. Whatever may be the magnet's inertial motion, an observer moving with it will see a pure magnetic field; an observer in another state of inertial motion will see a mixture of magnetic and electric fields. That is just what moving magnetic fields look like, Einstein supposed—just as, in the later special theory of relativity, observers see moving clocks slow and rods shrink, while co-moving observers do not.

4.2 Field transformations and the relativity of simultaneity

With this notion of field transformations, Einstein had created a potent device and it remained of central importance. For it was how Einstein would finally show in 1905 that Maxwell's electrodynamics conformed to the principle of relativity after all. The difficulty Einstein faced, however, was that no fully relativistic formulation of Maxwell's electrodynamics was possible just using this new device of field transformations. It had to be coupled with the novel account of space and time in special relativity. A simple thought experiment—*not* due to Einstein—shows that the device of field transformations requires Einstein's later notion of the relativity of simultaneity if it is to be implemented in a relativized Maxwell's theory.[11]

Consider a very long coil of wire with a rectangular cross section. When a current is passed through the coil, a uniform magnetic field **H** appears inside, with the magnetic field running along the axis of the coil. The wire consists of a lattice of immobile positive charges, with the current due to the motion of negatively charged electrons. The density of positive and negative charges will balance exactly so the wire carries no net charge. A section through the coil is shown in Figure 10, as it is seen by the "co-moving observer," an observer who moves with the coil.

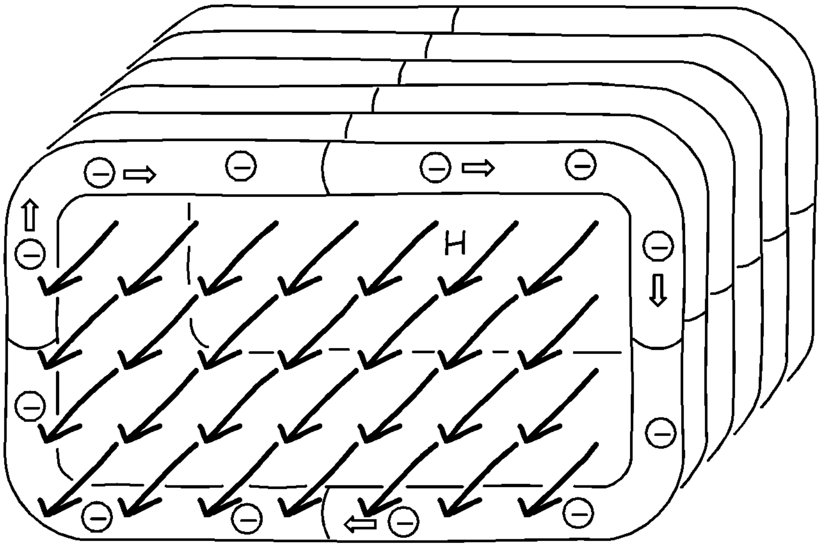

Figure 10. Magnetic field inside a coil as seen by a co-moving observer

We now set the coil into uniform motion. Figure 11 shows how it will appear to a "resting observer," that is, one who remains at rest while the magnet moves past. Following Einstein's prescription, a resting observer will see an induced electric field **E** associated with the magnetic field. From the case of the magnet and conductor (Figure 9 above) we can see that the induced electric field will be perpendicular to both the magnetic field and the direction of motion. Since the magnetic field is uniform, the induced electric field will be uniform as well and it will run from the bottom of the coil to the top. Since there is no magnetic field outside an infinitely long coil, the electric field lines of force will terminate in the wire. Maxwell's theory is clear on what that means: electrical lines of force can only terminate in charges. The result is that the top of the coil carries a net negative charge and the bottom a net positive charge, both of which are not seen by the co-moving observer.

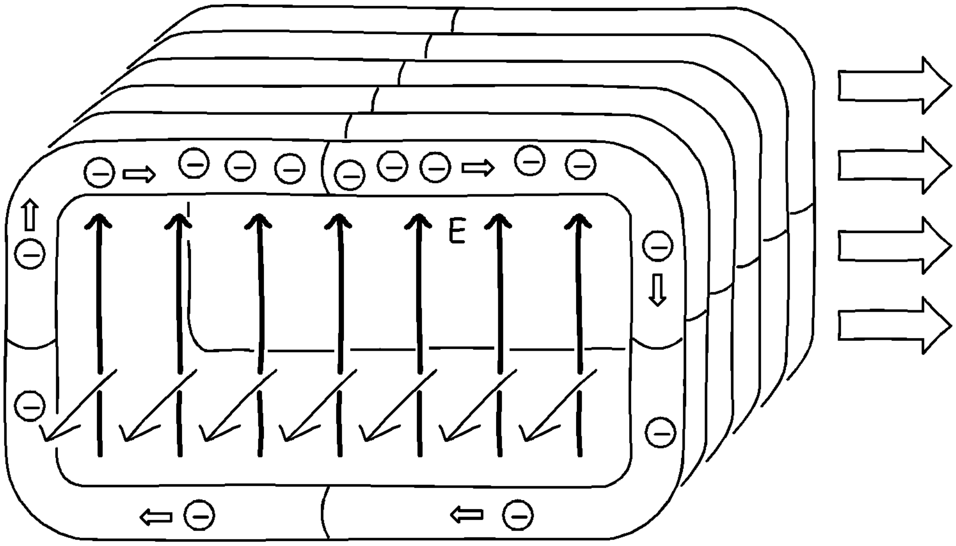

Figure 11. Induced electric field inside a moving coil

How can this happen? The co-moving observer judges the current carrying electrons to take the take the same time to move from left to right as from right to left. The resting observer does not. They take more time to traverse the coil in the left to right direction (with the motion) and less in the other direction (against the motion). As a result, there is an accumulation of negative charges on the top and dilution of negative charges on the bottom of the coil, yielding net negative and positive charges respectively. [12]

This difference of traversal times cannot happen in classical (Galilean) kinematics. If one observer judges the two traversal times to be equal, then so must all observers. The difference can arise in special relativistic kinematics; it is a direct expression of the relativity of simultaneity. We have already seen it above for light signals. The platform observer of Figure 1 judges the traversal times of light signals over the platform to be the same in both directions. The moving observer judges the traversal times to conform to Figure 2; they are not the same. This disagreement immediately leads to their differing judgments concerning the simultaneity of the events at A and B; that is, to the relativity of simultaneity.

This thought experiment shows that sufficient pursuit of Einstein's device of field transformations in Maxwell's electrodynamics must eventually force the relativity of simultaneity. The device cannot be used satisfactorily for the realizing of the principle of relativity until Einstein adopts the novel account of space and time of special relativity. The thought experiment is not Einstein's. We do not know the precise path that Einstein took from these field transformations to the space and time transformations of special relativity. He may have used physical reasoning such as in the thought experiment. Or he may have arrived at the result by mathematical analysis of the formal properties of Maxwell's equation, much as we might imagine Lorentz doing. Or he may have used both. We do know, however, that it took years and that several other considerations entered.

4.3 Einstein considers an emission theory of light

Einstein could not see how to formulate a fully relativistic electrodynamics merely using his new device of field transformations. So he considered the possibility of modifying Maxwell's electrodynamics in order to bring it into accord with an emission theory of light, such as Newton had originally conceived. There was some inevitability in these attempts, as long as he held to classical (Galilean) kinematics. Imagine that some emitter sends out a light beam at c. According to this kinematics, an observer who moves past at v in the opposite direction, will see the emitter moving at v and the light emitted at $c+v$. This last fact is the defining characteristic of an emission theory of light: the velocity of the emitter is added vectorially to the velocity of light emitted.

Einstein ran into numerous difficulties in his explorations of an emission theory. The principle difficulty, however, was this: if the emission theory was to be formulated as a field theory in which light is fully described as a propagating wave, then a light wave must somehow encode within it the velocity of its emitter, so that the theory could assign the correct velocity of propagation to each wave. No such encoding seemed possible, however, since experience showed that light waves were fully characterized simply by their intensity, color and polarization.

That no field theory can do this is not immediately obvious. My conjecture (Norton, 2003, §§5-6) is that Einstein's objections to an emission theory of light can be made transparent through a celebrated thought experiment that he first hit upon at the age of 16 and whose continuing cogency for Einstein would otherwise be unclear. As reported in his *Autobiographical Notes* (1949, p. 49-50) and elsewhere, he imagined chasing a beam of light at *c*. The result would be the observing of an electromagnetic waveform, frozen in space. "There seems to be no such thing, however," Einstein retorted, "neither on the basis of experience nor according to Maxwell's equations." Yet the retort is untroubling to an ether theorist. Maxwell's equations *do* entail quite directly that the observer would find a frozen waveform; and the ether theorist does not expect frozen waveforms in our experience since we do not move at the velocity of light in the ether. Why, then, was the thought experiment singled out for special attention in Einstein's recollections if its cogency is so doubtful?

The cogency becomes apparent if we place the thought experiment in Einstein's investigations of an emission theory of light. According to an emission theory, we should find frozen or slowed light waveforms if there are any sources of light moving sufficiently rapidly with respect to us. But we don't—just as Einstein remarked in his thought experiment. Even if we don't find them, the possibility of these static waveforms must be admitted by an emission theory if it is also a field theory. Now the sorts of static electric and magnetic fields possible were then well understood. Their investigation involved none of the relativistic complications of motion, rapid or otherwise. So an emission theory would have to agree with then current theories of electrostatics and magnetostatics, as Maxwell's theory did. In agreement with these theories, Maxwell's theory prohibits frozen waveforms--just as Einstein remarked in his thought experiment—and so also should a viable emission theory.

Finally a field theory, patterned even loosely after Maxwell's theory, will use the present state of a wave to determine its velocity of propagation. This is what allows field theories to be deterministic, so that according to them the present can determine the future. Yet just such determination is denied by an emission theory if it is also a field theory. If the present state of a light wave is determined fully by its intensity, color and polarization, it can have any velocity of propagation. As Einstein's thought experiment shows, it is even possible to have the extreme case of a completely frozen wave with no velocity of propagation; we merely need to move an observer at *c* with respect to the light's source. If an emission theory can be formulated as a field theory, it would seem to be unable to determine the future course of processes from their state in the present. As long as Einstein expected a viable theory of light, electricity and magnetism to be a field theory, these sorts of objections would render an emission theory of light inadmissible.

4.4 Return to Maxwell's theory

The early fruitlessness of Einstein's device of field transformations and his failed attempts to modify Maxwell's theory are just two episodes extracted from nearly a decade of thought on the problem of relative motion in electrodynamics. That thought must also have been entangled with his other investigations of what would become the light quantum hypothesis and the associated ebbing of his confidence in the exact validity of Maxwell's theory. Einstein recalled his reaction to these doubts and failures in his *Autobiographical Notes* (1949, p.49):

> Gradually I despaired of the possibility of discovering the true laws by means of constructive efforts based on known facts. The longer and more desperately I tried, the more I came to the conviction that only the discovery of a universal formal principle could lead us to assured results.

So he sought a theory that merely restricted the possibilities by means of principles whose grounding was secure. That decision brought special relativity to us as a theory founded on two postulates. In the light postulate, Einstein recorded one thing of which he had become sure. An emission theory fails. As he wrote in his 1905 paper "light is always propagated in empty space with a definite velocity *c which is independent of the state of motion of the emitting body*."

As the walls closed in, Einstein was brought to his final crisis. In a story that has been often told (e.g. Stachel, 2002, p. 185), Einstein visited his friend Michele Besso some five or six weeks prior to the completion of the 1905 paper, bringing his struggle with him. The next day, he reported with glee to his friend that he had found the solution, the relativity of simultaneity. He recalled in his *Autobiographical Notes* (1949, p. 51) how his analysis had been decisively furthered by reading the philosophical writings of David Hume and Ernst Mach. While Einstein did not elaborate on how they assisted him, it is not hard to guess. Both Hume and Mach stress that concepts are only warranted in so far as they are anchored in experience. Einstein now saw that the classical notion of time incorporated a concept of absolute simultaneity that had no basis in experience. Emboldened by Hume and Mach's critiques, Einstein discarded the classical notion and the path to the completed theory was opened.

4.5 Stellar Aberration

The analysis of stellar aberration provides a simple illustration of the different theories of light and their associated kinematics. It also supplies one of the most direct expressions of the relativity of simultaneity in observables. Indeed the expression is so direct that I shall also suggest that it may have been important in the closing stages of Einstein's reflections.

In 1727, James Bradley observed that the motion of the earth around the sun affected the direction of starlight arriving at the earth. The simple prescription for computing the change of direction is shown in Figure 12.[13]

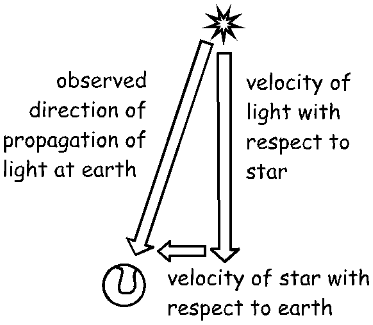

Figure 12. Stellar aberration

The velocity of the light with respect to its emitter, the star, is added vectorially to the velocity of the star with respect to the earth. The *direction* of the resulting compounded motion is the direction of the starlight observed on earth. If this vectorial addition gave the correct direction, how could we avoid concluding that it also gave the correct velocity? To conclude that would be to accept an emission theory of light.

The passage to the emission theory is so natural that one might wonder how an ether-based, wave theory of light could possibly accommodate Bradley's result. Yet it turns out to be quite easy, as is shown in Figure 13. While the wave fronts of light propagating from a star are spherical, the small portion of the wave fronts reaching the earth from a very distant star are virtually flat, so they become plane waves as depicted in the figure.

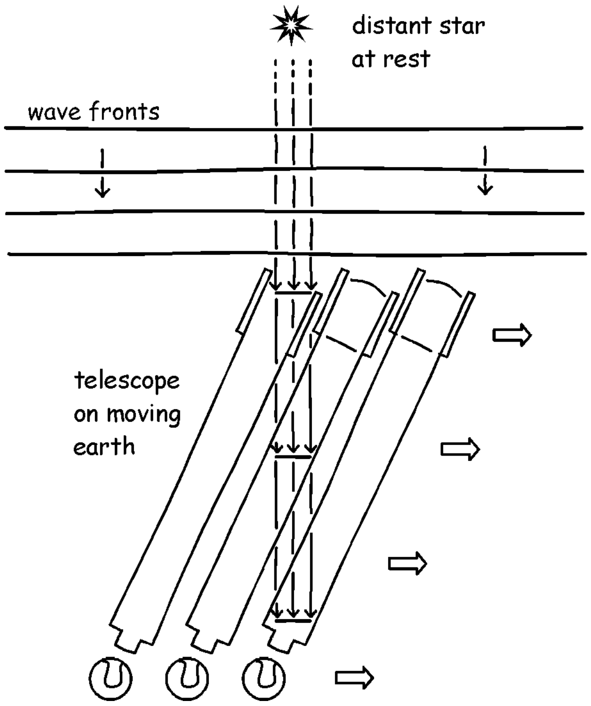

Figure 13. Stellar aberration in an ether-based, wave theory of light

The wave fronts propagate towards a telescope on the earth that moves from left to right. The telescope must be tilted as shown if a wave front that enters the front of the telescope is to pass along the barrel of the telescope to the observer's eyepiece. Otherwise the trailing telescope wall will intercept the wave before it reaches the eyepiece. The tilting of the telescope alters the apparent direction of the starlight in just the amount of Bradley's result.

This successful accommodation of aberration to the wave theory appears to fail completely, however, if we also demand that the wave theory respect the principle of relativity. For now we should expect the same observable result if we conceive the star as at rest and the earth moving (as in Figure 13); or if we conceive the star moving and the earth at rest. According to the principle of relativity, the effect should only depend on the relative velocity of earth and star and not on which is conceived as moving. Using classical notions of space and time, we arrive at the second case of a resting earth by a Galilean transformation of the arrangement in Figure 13. The result is shown in Figure 14.

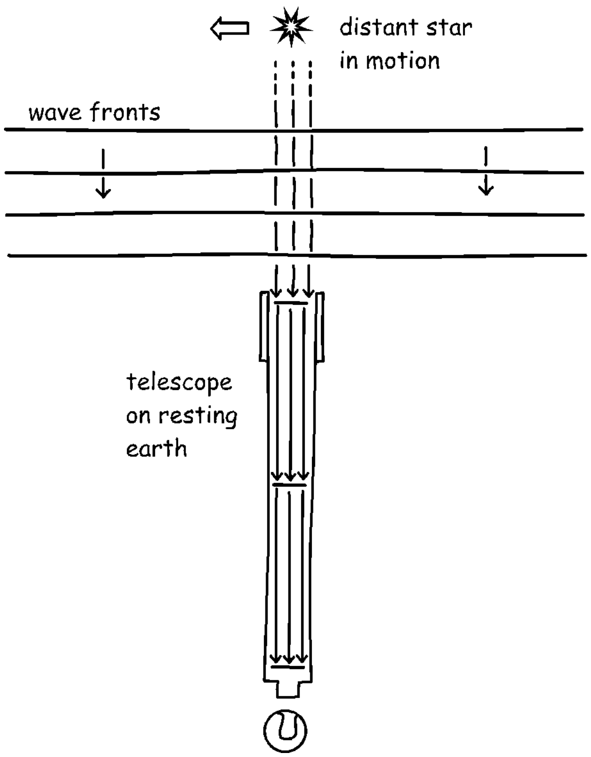

Figure 14. Galilean transformation of Figure 13 to a resting earth

The transformation brings the earth to rest and, at the same time, sets the star in motion in the opposite direction. The wave fronts remain perpendicular to the line joining the star and the earth. With this arrangement, it is immediately apparent that the effect of stellar aberration is obliterated. The motion of the star no longer has an effect on how we must aim the telescope on earth. If the telescope is pointed directly at the star, its light will pass to the eyepiece.[14] We seem to have a violation of the principle of relativity; whether the earth or the star moves absolutely can be determined by checking for the presence or absence of stellar aberration.

One of the great achievements of Lorentz's 1895 *Versuch* was to show that, in Maxwell's electrodynamical theory of light, stellar aberration does depend solely on the relative velocity after all, or at least to the first order quantities accessible to measurement. His demonstration depended upon the theorem of corresponding states. It requires us to use a Lorentz transformation, not a Galilean transformation, if we want to infer from the arrangement of Figure 13 how light would propagate were the earth at rest. The effect of applying a Lorentz transformation is shown in Figure 15.

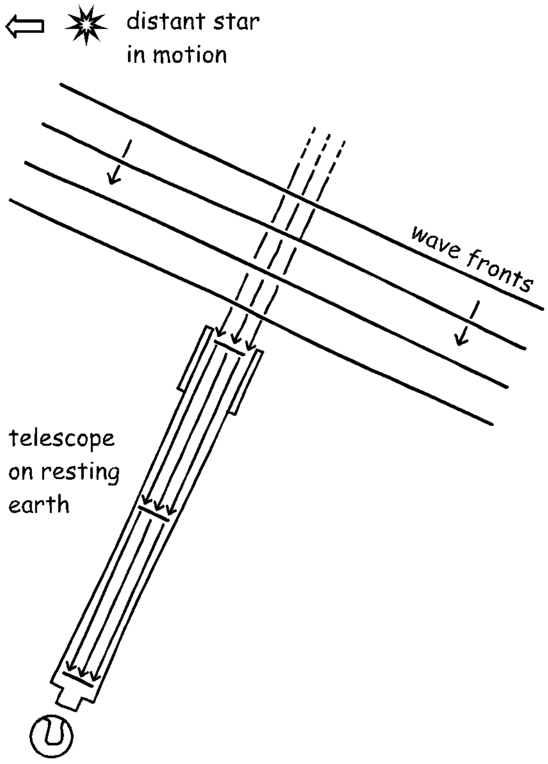

Figure 15. Lorentz transformation of Figure 13 to a resting earth

As before, the earth is brought to rest and the star is set in motion in the opposite direction. In addition, the Lorentz transformation rotates the wavefronts so that they are no longer perpendicular to the line connecting the star and the earth. That means that the telescope on earth must be directed away from that line if the light from the star is to reach the eyepiece. The relative motion of the earth and the star once again affects the apparent direction of propagation of the starlight just by the amount of Bradley's result. The observed result of stellar aberration is recovered and the outcome is in conformity with the principle of relativity. In the older analysis, the direction of propagation of a plane wave would be perpendicular to the wave fronts only for an observer at rest in the ether; for all others it would fail. Earth-bound astronomers, observing starlight, could use this failure to establish their absolute motion. Lorentz' analysis deprives them of this possibility, since, in his analysis, the direction of propagation is perpendicular to the wave fronts for all inertial observers.

What is intriguing about the core effect, the rotation of the wave fronts of Figure 15, is that it is due *entirely* to Lorentz's local time. Formally it is exactly the same effect as the rotation of the platform due to the relativity of simultaneity shown in Figure 4 above. That the vertically moving platform or wave fronts in the figures are oriented horizontally depends on a judgment of simultaneity. If observers change their state of motion and thus their judgments of simultaneity, the immediate outcome is the rotations shown in Figures 4 and 15.

Lorentz gives a very different interpretation of this effect than does Einstein's special theory of relativity. For Lorentz, it is simply a matter of consulting Maxwell's electrodynamics to find the effect of a star's motion on the light it emits. Such consultation is made through the theorem of corresponding states and, in this case, we find that the effect is a rotation of the wave fronts. For Einstein's special theory, the effect has nothing in particular to do with electrodynamics, but comes directly from space and time. It arises whenever a long object, be it a long platform as in Figures 3 and 4 or a wave front as in Figure 15, moves perpendicular to its length and we change the state of motion of the observer.

I'd like to suggest that this consideration might have been important to Einstein in the closing stages of the reflections leading to his 1905 paper. It might have provided a way to see how the relativity of simultaneity was grounded in experience, just as Hume and Mach had demanded of our scientific concepts. In the early 1950s, Robert Shankland visited Einstein with the express purpose of learning of the degree to which the Michelson-Morley experiment had influenced Einstein. To his dismay, he found Einstein barely able to recall whether he even knew of the experiment prior to his 1905 paper and suggesting that, if he did, he took the result for granted. That reaction is not hard to understand. The null outcome of the experiment is a direct result of the principle of relativity and the experiment cannot decide between an emission theory of light or one based on the light postulate. Instead, Einstein volunteered, he had been more influenced by Fizeau's measurement of the speed of light in moving water and by stellar aberration. Shankland, unfortunately, did not ask Einstein to explain. We cannot be sure of what Einstein intended with these remarks. He may merely have meant that these experiments allowed a decision between a resting ether and a fully dragged ether. I lean towards another far more interesting explanation.

Einstein, I propose, interested in realizing a principle of relativity and in anchoring his theorizing in facts of experience, would have attended closely to what experience delivered. The earth moves to new frames of reference as it slowly changes its direction of motion in its orbit around the sun. In the course of a year, earth-bound astronomers sample the light from a given star in many different frames of reference. They find the direction of deflection of stellar aberration and the associated rotation of the wave fronts to be different in each frame. So Einstein could take these views and read from them the transformation between the associated frames of reference that must figure in his principle of relativity. It must be one that can rotate vertically moving objects such as in Figures 3 and 4. The result would be a transformation that employs local time, that is, that embodies the sort of dislocation of temporal parts of the relativity of simultaneity. It could not be a Galilean transformation since that transformation does not rotate wave fronts.

What is important is that this dislocation of simultaneity would be read *from the observations* of stellar aberration and would be independent of Maxwell's electrodynamics and Lorentz's theorem of corresponding states. The only assumptions would be that starlight is a propagating waveform conforming to the principle of relativity.[15] If our concept of time was to be grounded in experience, here was experience calling for a concept of time that incorporated the relativity of simultaneity.

5. $E=mc^2$

5.1 The result

Shortly after completing his paper on special relativity, Einstein found another consequence of the theory that he described in a short note "Does the Inertia of a Body Depend upon its Energy Content?". The basic notion was as simple as it was profound. Any quantity of energy, an amount "E" for example, also carries a mass "m" in direct proportion to the energy. The mass is computed by dividing the energy E by the number c^2. That number is so large that the associated mass is usually very tiny. Conversely, any mass m is also a quantity of energy E, where the conversion is effected by multiplying m by c^2. Because c^2 is so large, even a very small mass is associated with an enormous amount of energy.

This result of the inertia of energy can be applied whenever mass or energy transforms. Sometimes the effect is an imperceptible curiosity. When we talk on a battery powered cell phone, the battery loses energy as it powers the phone. The accompanying, miniscule loss of mass of the battery is imperceptible to us. On other occasions the effect is world changing. When Uranium-235 undergoes fission, it breaks into other elements whose total mass turns out to be slightly less than that of the original Uranium. That slight mass deficit manifests as an enormous quantity of energy in heat and radiation. As was discovered decades later, that process can power atom bombs or nuclear power plants.

To the casual reader, virtually all of Einstein's demonstrations of $E=mc^2$ seem curiously complicated, drawing on arcane results in electrodynamics, now generally regarded as more obscure than the result to be shown. Even a mid-century derivation (Einstein, 1946), offered as especially simple, takes the pressure of radiation as a primitive notion. The reasons for this obliqueness lie in the physics and in its history. Special relativity, as a theory of space and time, cannot make pronouncements by itself on energy, mass and matter. It can only constrain the ways that they can manifest in space and time: they must be governed by laws that admit no absolute velocities. So some extra physical assumption must be supplied to determine which of the possibilities is realized. In Einstein's case, that extra assumption is conveyed by electrodynamics. The choice of electrodynamics for this purpose is entirely natural. The inertia of energy is a result already to be found in Maxwell's electrodynamics, just as the kinematics of special relativity were first discovered in Maxwell's theory. The real import of Einstein's demonstrations is to show that the inertia of energy cannot be localized to electrodynamics alone. Once it is secured there, relativity theory demands that it must hold for all forms of energy.[16]

5.2 A demonstration

The following is a version of Einstein's (1905a) demonstration, simplified along the lines of Einstein (1946).[17] It is designed to show that if the inertia of energy is realized in Maxwell's electrodynamics, it must be realized for all forms of mass and energy. The inertia of energy is expressed in Maxwell's theory for unidirectional radiation as follows: a quantity of radiant energy E carries momentum E/c in the direction of its motion. (To make the result familiar, assume that momentum has magnitude mc where m is the mass of radiation and we have $E/c=mc$ so that $E=mc^2$.)

A body with mass m' at rest emits two quantities of radiant energy $E'/2$ in opposite directions, as shown in Figure 16.

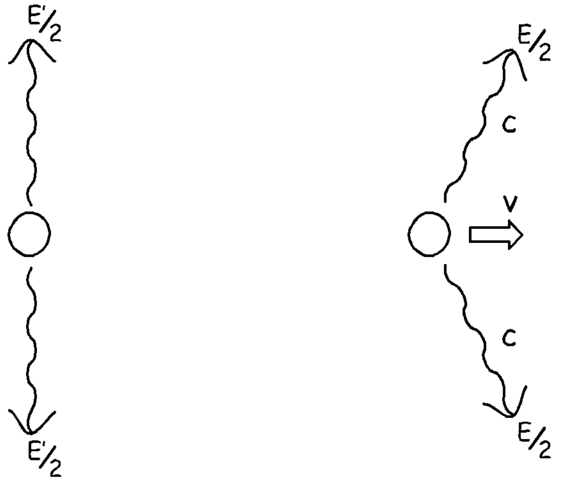

Figure 16. A mass emits two quantities of radiation

Because of the symmetry of the emission, the body remains at rest. We now view the process from a frame in which the body moves perpendicularly to the direction of emission at v, and in which it has mass m. The quantities of radiant energy are now $E/2$

in the new frame. Thus they carry momentum $E/2c$ in the direction of propagation and a portion of that momentum in the ratio v/c lies in the direction of the body's motion. That portion is $(E/2c)(v/c) = (1/2)(E/c^2)v$. The law of conservation of momentum tells us that momentum gained by the radiation must equal that lost by the body. In the direction of the body's motion, the radiation has gained momentum $2\text{x}(1/2)(E/c^2)v = (E/c^2)v$. So the momentum of the body must be reduced by the same amount. The momentum of the body is mv and it must reduce by $(E/c^2)v$ as a result of the emission. Since the emission did not accelerate the body in its rest frame, the same will be true in this frame. Therefore the body's velocity remains v. So the decrease of momentum must come from a reduction in the mass m of the body. In sum, the body loses energy E and, as a result, loses momentum $(E/c^2)v$, which corresponds to a loss of mass of (E/c^2). This is the inertia of energy, now demonstrated for any body whatever that can emit radiation in the way shown.

6. Conclusion

> "The special theory of relativity owes its origins to Maxwell's equations of the electromagnetic field."
> Albert Einstein, *Autobiographical Notes*, p. 59

In our brief review of the origins of Einstein's theory, we have seen much to affirm Einstein's judgment. The theory was already implicit in Maxwell's electrodynamics—so much so that Lorentz was able to discover its essential mathematical structure without realizing that he had chanced upon a new theory of space and time. On the basis of this theory, Henri Poincaré had also begun to speak of the principle of relativity as one of the principles to which all physics must be subject. In an analysis that has excited and maddened later commentators, Poincaré interpreted Lorentz's local time in terms of the synchronizing of clocks by light signals, just as Einstein did later, but without conveying a sense that this construction was the core of a new physical theory of space and time. (See Darrigol, 1995.)

In assessing both Lorentz's and Poincaré's work, one must guard against interpreting their thought and goals solely in terms of their proximity to Einstein's work. It is entirely possible to recognize that no experiment will reveal the earth's motion through the ether and even to codify this expectation as a principle of relativity, without demanding that our theories be overturned so as to eradicate all trace of an ether and its state of rest. Another principle of physics, also discussed by Poincaré, illustrates this. The second law of thermodynamics assures us of an inexorable unidirectionality in time for all thermal processes and we routinely use it, while fully recognizing that its unidirectionality need not be reflected in the fundamental physics that governs thermal processes. The fundamental physics hides its time reversibility systematically from experimental detection, just as we might imagine that the physics of electricity and magnetism systematically hides the ether state of rest from us.[18] Poincaré's seemed not to have regarded his analysis of Lorentz's local time as a physical discovery. Rather it illustrated a persistent theme of his thought, the conventional character of parts of our science. Among many systems for synchronizing clocks we choose the one that we find most convenient, in that it makes the expression of laws of physics most simple. The choice is not imposed by nature but by our preferences and the point is illustrated by an odd synchrony scheme that gives simple expression for results in electrodynamics. (See Poincaré, 1898, p. 222.)

Einstein's approach was quite different, characterized by an enduring conviction that the principle of relativity had to be realized throughout electrodynamics, even when years of investigation seemed to show the goal unrealistic and unachievable. While electromagnetism could not reveal the ether state of rest in the context of the magnet and conductor thought experiment, as Föppl had already pointed out, there were other equally simple experiments in electrodynamics that would, or so it seemed. The device Einstein learned from the thought experiment, field transformations, would have proved infertile as a means for realizing the principle of relativity. A thorough examination of its use in Maxwell's theory would have shown that different parts of the theory require different field transformations, as long as the kinematics remained Galilean (See Norton, 2003, §2). Undaunted, Einstein was willing to sacrifice the greatest success of 19th century physics, Maxwell's theory, seeking to replace it by one conforming to an emission theory of light, as the classical, Galilean kinematics demanded. With the failures mounting and his options exhausted, Einstein would entertain an extraordinary and desperate thought. Could he realize the principle of relativity in electrodynamics if he reshaped the very notion of time itself? This final gambit succeeded. When such stubbornness prevails, we wonder at its prescience; if it fails, we lament its folly.

Einstein is inseparably linked with analyses dense in light signals and the clocks they synchronize. What is notable in the above account is how little they figured in Einstein's path to special relativity. They were decisive in the final moments, some five to six weeks prior to the completion of the theory, when Einstein probably used them in his last, desperate gambit. But there is no evidence in the long years of investigation preceding that Einstein gave any serious thought to light signals and clocks. He did ponder light as it is judged by observers in different states of motion. But that was light as a propagating waveform in Maxwell's electrodynamics, not light as a signal, a point moving at *c*. The light of his original thought experiment at age 16 was light as a propagating waveform, for he immediately recoiled at the resulting temporally frozen waveform. The optical experiments that Einstein singled out as important in his thought prior to the 1905 paper were stellar aberration and Fizeau's experiment. Both admit a very simple analysis in which the relativistic rule of velocity addition is applied to light signals. However he seemed not to conceive them that way in 1905. His 1905 analysis of stellar aberration is given in terms of transforming light as propagating waveform, much as in Section 4.5 above; and, in 1907, Einstein reported that he had only then learned from Laue of this perspicacious analysis of Fizeau's experiment.

The dominance of light signals and clock synchrony seems to be very much an artifact of Einstein's own final steps, their undeniable pedagogic value and of our own preoccupation with them.[19] They are not necessary to special relativity or to the relativity of simultaneity. They need not appear at all in a spacetime formulation of special relativity, where the relativity of simultaneity arises naturally as our freedom to slice up the spacetime into spaces in many different but equivalent ways.

References

Born, Max (1962) *Einstein's Theory of Relativity*. New York: Dover.

Darrigol, Olivier (1995) "Henri Poincaré's Criticism of *Fin De Siècle* Electrodynamics," *Studies in History and Philosophy of Modern Physics*, *26*, pp. 1-44.

Darrigol, Olivier (2000) *Electrodynamics from Ampère to Einstein*. Oxford: Oxford University Press.

Earman, John, Glymour, Clark and Rynasiewicz, Robert (1983) "On Writing the History of Special Relativity," *PSA 1982*, Volume 2. Peter D. Asquith and Thomas Nickles, eds., East Lansing: Philosophy of Science Association, pp. 403-416.

Einstein, Albert (1905) "Zur Electrodynamik bewegter Körper," *Annalen der Physik*, *17*, pp. 891-921. ("On the Electrodynamics of Moving Bodies.")

Einstein, Albert (1905a) "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" *Annalen der Physik*, *18*, pp. 639-641. ("Does the Inertia of a Body Depend upon its Energy Content?")

Einstein, Albert (1917) *Über die spezielle and die allgemeine Relativitätstheorie (Gemeinverständlich)*. Braunschweig: Friedr. Vieweg & Sohn; 15th expanded edition, trans. R. W. Lawson, *Relativity: the Special and the General Theory.* London: Methuen, 1954.

Einstein, Albert (1946) "An Elementary Derivation of the Equivalence of Mass and Energy," pp. 116-19 in *Out of my Later Years*. New York: Bonanza, 1990.

Einstein, Albert (1949) *Autobiographical Notes*. P. A. Schilpp, trans. and ed., La Salle and Chicago: Open court, 1979. (Note to editors: this is a separate printing with pagination different from the full Schilpp volume.)

Föppl, August (1894) *Einführung in die Maxwell'sche Theorie der Elektricität*. Leipzig: B. G. Tuebner.

Galison, Peter (2003) *Einstein's Clocks, Poincaré's Maps: Empires of Time.* New York: W. W. Norton.

Holton, Gerald (1969) "Einstein, Michelson and the 'Crucial' Experiment," *Isis*, *60*, pp. 133-69; Ch. 9 in *Thematic Origins of Scientific Thought: Kepler to Einstein.* Cambridge, MA: Harvard University Press, 1973.

Janssen, Michel (1995) *A Comparison Between Lorentz's Ether Theory and Special Relativity in the Light of the Experiments of Trouton and Noble.* PhD Dissertation. Department of History and Philosophy of Science, University of Pittsburgh.

Janssen, Michel (2002) "Reconsidering a Scientific Revolution: The Case of Einstein versus Lorentz," *Physics in Perspective*, *4*, pp. 421–446.

Janssen, Michel (2002a) " COI Stories: Explanation and Evidence in the History of Science," *Perspectives on Science*, *10*, pp. 457 – 522.

Janssen, Michel (2003) "the Trouton Experiment, $E=MC^2$, and a Slice of Minkowski Spacetime," pp. 27-54 in A. Ashtekar *et al.* (eds.) *Revisiting the Foundations of Relativistic Physics*. Dordrect: Kluwer.

Lorentz, Hendrik A. (1895) *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden: E. J. Brill. ("Draft of a theory of electrical and optical phenomena in moving bodies.")

Lorentz, Hendrik A. (1904) "Electromagnetic Phenomena in a System Moving with any Velocity less than that of Light," *Proceedings of the Academy of Sciences, Amsterdam*, *6*; reprinted pp. 11-34 H. A. Lorentz *et al.*, *The Principle of Relativity.* Trans W. Perrett and G. B. Jeffrey. Methuen, 1923; Dover, 1952.

Miller, Arthur (1981) *Albert Einstein's Special Theory of Relativity*. Reading, MA: Addison-Wesley.

Norton, John D. (2004) "Einstein's investigations of Galilean covariant electrodynamics prior to 1905," *Archive for History of Exact Sciences*, **59**, pp. 45-105.

Poincaré, Henri (1898) "The Measure of Time," in *The Value of Science: Essential Writings of Henri Poincaré*. New York: Modern Library, 2001.

Rynasiewicz, Robert (2000) "The Construction of the Special Theory: Some Queries and Considerations," in Howard and Stachel (2000), pp. 159-201.

Shankland, R. S. (1963/73) "Conversations with Einstein," *American Journal of Physics, 31*(1963), pp. 47-57; 41(1973), pp. 895-901.

Stachel, John (1987) "Einstein and Ether Drift Experiments," *Physics Today*, **40**(1987), pp. 45-47; reprinted Stachel (2002), pp. 171-76.

Stachel, John (2002) *Einstein from 'B' to 'Z.' : Einstein Studies, Volume 9*. Boston: Birkhäuser.

Stachel, John *et al.* (eds.) (1989) *The Collected Papers of Albert Einstein: Volume 2: The Swiss Years: Writing, 1900-1902.* Princeton: Princeton University Press. ("*Papers*, Vol. 2.")

Stachel, John *et al.* (1989a) "Einstein on the Theory of Relativity," Headnote in *Papers*, Vol. 2., pp. 253-74.

Whittaker, Edmund T. (1951) *A History of the Theories of Aether and Electricity*. London: Thomas Nelson & sons, 1951, 1953; reprinted Dover, 1989.

---

[1] I am grateful to Tony Duncan, Allen Janis, Michel Janssen and Robert Rynasiewicz for helpful comments.

[2] See Holton, 1969, and for discussion informed by more recent discoveries in Einstein's correspondence, Stachel, 1987.

[3] Einstein wrote in his *Autobiographical Notes* (1949, p. 31): "Newton, forgive me: you found just about the only way possible in your age for a man of highest reasoning and creative power."

[4] Einstein (1917, p.41) wrote that special relativity had been "developed from electrodynamics as an astoundingly simple combination and generalization of the hypotheses, formerly independent of each other, on which electrodynamics was built."

[5] A frame of reference is a system for assigning positions and times to events in association with a particular state of motion. It is conveniently realized by imagining space filled with a lattice of sticks and that every point of the lattice is equipped with a synchronized clock. Coordinates of time and space are assigned to events by a numbering system for the points of the lattice and by the readings of the clocks. An inertial motion is a uniform, straight-line motion naturally adopted by masses moving free of net forces. An inertial frame of reference is one that moves inertially.

[6] For a history of ether theories and electrodynamics in the 19th century, see Whittaker (1951) and Darrigol (2000); and, for a treatment more narrowly focused on Einstein's 1905 paper, Miller (1981).

[7] These experiments were expected to produce a measurable effect that is a function of v/c, where v is the speed of the earth. Writing the effect as a power series, it is Effect = $A(v/c) + B(v/c)^2 + C(v/c)^3 + \ldots$ A first order experiment seeks to measure $A(v/c)$.

If that is zero, a second order experiment would try to measure $B(v/c)^2$. Since (v/c) is very small, terms in $B(v/c)^2$ are still smaller and extraordinarily difficult to measure.

[8] For a thorough discussion, see Janssen (1995).

[9] This section draws on work of recent decades that had profited from extensive scrutiny of material in the Einstein archive. See Stachel *et al*. (1989), Stachel (2002, Part IV), Rynasiewicz (2000) and Earman *at al*. (1983). My own attempts to extend these accounts is Norton (2003), which expands on many of the points made in the text.

[10] The version I will develop is a slight variant form given by August Föppl (1894, pp. 309-10) in an electrodynamics text which Einstein probably read.

[11] A simpler argument is that Maxwell's theory entails the constancy of the speed of light and that constancy, along with the principle of relativity, entails the relativity of simultaneity. This simpler argument, however, does not reveal how the kinematics of special relativity must permeate through even the simplest electrodynamical processes if the principle of relativity is to be respected; showing that is the function of this magnet and coil thought experiment.

[12] An analogy: on a circular track, racing cars will accumulate on the slow side and dilute on the fast side.

[13] The figures that follow greatly exaggerate the change of direction. It is about 20 seconds of arc, which would be imperceptible in a properly scaled figure. The figures also show the special case of a star located in a direction perpendicular to the earth's motion.

[14] One might try to avoid the problem, as Born (1962, p. 141) suggests, by supposing that the direction of propagation is not perpendicular to the wave fronts. However this

---

might be achieved, it is not a solution available to someone, seeking to realize a principle of relativity. If the direction of propagation is perpendicular to the wave fronts in one frame, such as the one we designate as the ether frame, then, by the principle of relativity, that must also be true in any other inertial frame.

[15] Fizeau's experiment measured the Fresnel drag for light propagating in moving water. In a similar analysis, it can be seen to give experimental support to Lorentz's local time, independently of electrodynamical theory. See Norton (2003, §7).

[16] For another approach to this result, see Janssen (2003).

[17] Einstein's (1905a) derivation seems to have been complicated by an unfortunate definition of mass. His 1905 special relativity paper had defined mass as force/acceleration with the awkward outcome that mass has a different dependency on velocity according to whether the acceleration was parallel or transverse to the direction of its motion. Thus Einstein (1905a) demonstrated the inertia of energy for the rest mass only. The superior definition soon adopted by Einstein set mass equal to momentum/velocity and is used in the text.

[18] The analogy is not perfect, of course, for at that time it was recognized by many, Einstein and Poincaré included, that certain microscopic processes might reveal the hidden processes that make the second law of thermodynamics true only with high probability.

[19] For evidence of their enduring fascination, see Galison (2003).