432018 PHILOSOPHY OF PHYSICS (Spring 2002)

Lecture 13: The rise of space-time and the Special Theory of Relativity

Preliminary reading: Sklar, pp. 25-31.

In the next two lectures we are going to consider the *Special* Theory of Relativity. Specifically, in this lecture we are going to introduce the concept of a *space-time* in a 'classical' setting and examine the historical background to this theory. Then, having done this, we will be in a position to look at the theory itself in the next lecture.

NOTE: The details of the different 'classical' space-times discussed in Section 1 and the mathematics outlined in Section 2 are not essential for this course. But, you should try to understand the basic philosophical and physical points that are being made.

1 'Classical' four-dimensional space-times

Having familiarised ourselves with the absolutist and relationist views of space and time, you may be tempted to think that they would simply 'fit together' to give us absolutist and relationist views of 'space-time' respectively. However, we shall see that the move from 'space and time' to 'space-time', allows us to distinguish several intermediate 'kinds' of space-time. To illustrate this, we shall start by describing how 'space-time' is, in some sense, a simple geometric structure which just 'combines' our three-dimensional view of space with a one-dimensional view of time to give us a four-dimensional space-time. We shall then see how we can distinguish several sorts of 'classical' space-time by varying the amount of additional structure which we impose on this simple idea of a space-time. In particular, although the details of this exercise are not important,[1] it will allow us to see how different space-times are compatible with certain aspects of the Newtonian theory of motion discussed in the earlier lectures.

1.1 The basic idea of space-time

In order to introduce you to the basic idea of a space-time, we shall motivate it by considering how the classical 'space and time' picture of the world can be 'translated' into a space-time picture. To see this, note that in Newtonian Mechanics, time is a *parameter* which allows us to distinguish states of affairs, say configurations of objects in our three-dimensional space, which occur at different times — see Figure 1 (a). But, in a space-time, we take time to be a further (i.e. a fourth) *dimension* and so we get a different picture of what is going on — see Figure 1 (b). The resulting geometric structure, the space-time, is often referred to as a four-dimensional *manifold.* And, this manifold is composed of space-time *events*, i.e. space-time *points*, which correspond in some ways to our earlier notion of 'positions at a time'. So, 'joining up the dots' corresponding to the space-time locations of a material body at intervening times, we get the *world-line* of that object — see Figure 1 (c) — where e and f (g and h) are events on the world-line of the particle p_1 (p_2).

Furthermore, in order to avoid any commitment to a notion of absolute space, we want to make sure that the 'axes' we use in our space-time diagrams aren't necessarily associated with an absolute frame of reference. To do this, we take a 'frame of reference' to be something that allows us to assign *coordinates* to an event. And, as such, the coordinates we assign to an event need not be absolute since they will depend on the frame of reference that we happen to adopt. For example,

- If you are working in a laboratory, you may want to specify the locations of objects by referring to their distance from the floor and two appropriate walls. You could find these distances by using a metre rule and once you have them you would be able to assign spatial coordinates to these objects relative to the walls and floor of the laboratory. If the objects were moving about, you could measure their locations in this way at different times. Once you have this information, you can use this 'space and time' information to construct a 'space-time' picture

[1] At least, for this course!

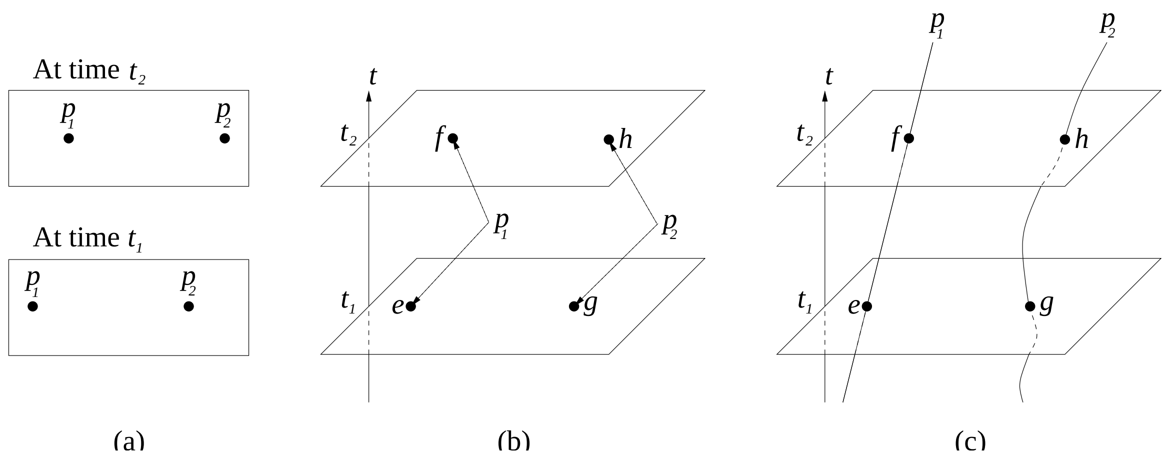

Figure 1: *The transition from the 'space and time' picture to the space-time picture. (a) we have two particles p_1 and p_2 which are at different locations at times t_1 and t_2. (b) The space-time picture corresponding to the 'space and time' picture given in (a). (c) By looking at the locations of p_1 and p_2 at other times we can trace out the world-line of these particles in the space-time picture. (Note: It is usual to suppress one of the spatial dimensions in such diagrams.)*

of what is happening in your laboratory. Obviously, if you were so inclined, you could extend this 'space-time' coordinate system (or frame of reference) defined within your laboratory to cover things which were happening outside of it.

- Alternatively, you may want to use a frame of reference that is fixed, not to your laboratory (which is at rest on the Earth), but to the Sun. As such, you could take your earlier laboratory measurements and then, by 'correcting' for the motion of the Earth around the Sun, you could translate them into coordinates which are defined by this new frame of reference. This will give us a different space-time picture that describes the same set of events.

Of course, using this idea many other frames of reference are possible, and these, in turn, will allow you to assign coordinates to events in many different ways.

1.2 Some different 'classical' space-times

Now, starting with the basic idea of space-time, we are going to create a number of new space-times by adding structure to them. And, in this context, those relations between events that are 'meaningful' in a given space-time will exhaust the spatiotemporal relations that are admissible within it. In particular, as well as allowing us to familiarise ourselves with frames of reference, the study of different space-times will allow us to see that we should not simply 'infer' facts about the structure of space-time itself from the coordinates being used.[2]

Indeed, formulating our theories using four-dimensional manifolds gives us greater flexibility. That is, there are now various different options available to us when we want to specify the nitty-gritty relations that may underpin our physical theories. To illustrate this, when we consider these different space-times (i.e. four-dimensional manifolds with different structures defined on them) we will see how they are related to one another. In particular, we shall see that the 'meaningfulness' or 'non-meaningfulness' of certain questions within a space-time can be a key factor in our assessment of the adequacy of a space-time as an 'arena' for a theory of motion.

Machian space-time

This is the 'sparsest' space-time that we will consider and it consists of a four-dimensional space-time with the following structure (see Figure 2):

[2]Moreover, to keep things simple, all of the space-times that we consider in this section are 'classical', i.e. they could be used as the background for a formulation of Newtonian mechanics, as opposed to (say) the relativistic space-times that we shall start to encounter in the next lecture.

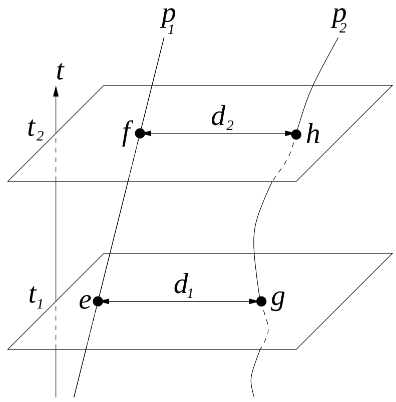

Figure 2: *Machian space-time: We can tell whether events are occurring at the same time, and if they are, we can tell how far apart they are (i.e. their relative [spatial] separation). But, we can't tell how much time has elapsed between two temporally separated events.*

- a [temporal] relation of absolute simultaneity, so we can determine whether events on particle world-lines are simultaneous (i.e. whether they occur at the same time). For example, the events e and g are simultaneous, as are the events f and h.
- an [instantaneous spatial] structure which is governed by Euclidean geometry. As such, spatial distances between simultaneous events are well defined. For example, the events e and g are separated by a spatial distance d_1, whereas the events f and h are separated by a distance d_2.

So, in this space-time, it is meaningful to ask questions like:

- 'Are events on particle world-lines simultaneous?'. But, if the answer is 'no' (e.g. e and h), it is *not* meaningful to ask: 'What is the spatial separation between these events?'.
- 'Is the spatial separation between two particles changing?'. But, if the answer is 'yes', it is *not* meaningful to ask: 'What is the *rate* of change of the distance?'.

As such, in Machian space-time, we can define a *rigid* reference frame to be a reference frame in which the spatial distance between any two points of the frame does not change with time. Indeed, these rigid reference frames allow us to assign coordinates to an event.

Note that, since we cannot talk about the rate at which the distance between two particles changes, we can't talk about things like the *speed* at which particles are moving towards (or away from) each other, and as such this space-time seems inadequate as a foundation for a theory of motion. We rectify this problem by adding some more structure to get the next space-time.

Leibnizian space-time

This is obtained from Machian space-time by adding a time metric, and as such we can now talk about the time *difference* between events that are on two different planes of absolute simultaneity (see Figure 3). In this space-time, it is meaningful to ask questions like:

- 'How fast is the [relative] distance between two particles changing?' and 'How fast is the *relative* speed of two particles changing?' *et cetera.* But, it is *not* meaningful to ask: 'How fast is this particle going?'.

As such, in this space-time, it is meaningful to ask questions about the *relative* motions of particles, but not about the *absolute* motions of particles. That is, by choosing a different rigid frame of reference, the situation illustrated in Figure 3 would look like the situation described in Figure 4 and there is no way of choosing between them. In particular, the key thing to notice here is that: in Figure 3 the world-line of p_1 is straight, whereas in Figure 4, the world-line of p_2 is straight.[3] Indeed, this appears to be the 'strongest' space-time that a relationist can utilise for his theory of motion.

[3]The significance of this will become apparent when we discuss neo-Newtonian space-time.

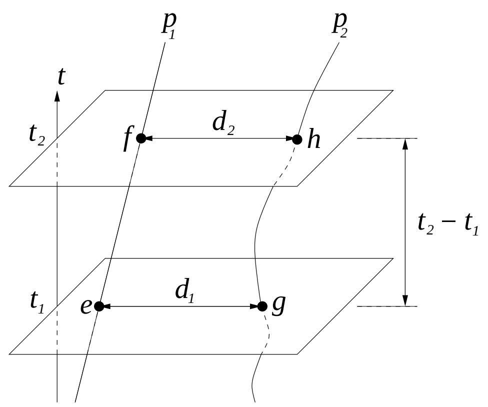

Figure 3: *Leibnizian space-time: We can now tell how much time has elapsed between two events and so we can now calculate relative velocities and accelerations.*

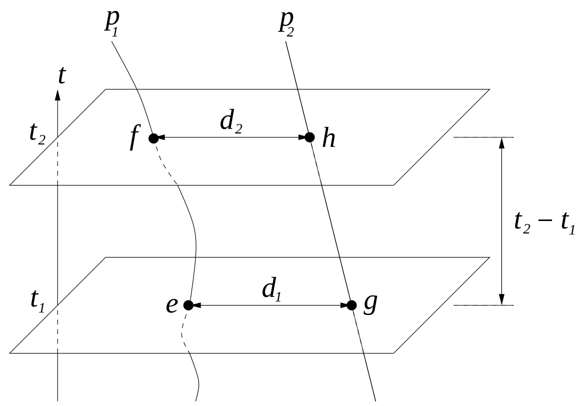

Figure 4: *In Leibnizian space-time, we have no way of choosing between the description of the events given in Figure 3 and the one given here (which utilises a different rigid frame of reference). In particular, we have no way of deciding whether the world-line of p_1 or p_2 is 'really' straight.*

But, recall that the 'bucket argument' raises a problem for the relationist, namely that the observed inertial effects couldn't be described in terms of relative motions. Specifically, for the absolutist, if we have two [extended] material objects, say one which is rotating relative to absolute space and one which isn't, then we only expect the former to experience inertial effects due to rotation. As such, we can't talk about such rotational effects in purely relational terms since, when described relationally, there is no way to account for why one of them is experiencing the inertial effects rather than the other.[4] So, although Leibnizian space-time would allow us to determine whether two [extended] material objects are rotating *relative* to one another, this isn't enough to account for such effects. The next space-time adds just enough structure to account for this.

Maxwellian space-time

Maxwellian space-time attempts to rectify this situation by adding a notion of absolute rotation, and as such, a way of deciding which bodies will experience the inertial effects of rotation and those that won't.

To be specific, Maxwellian space-time is obtained from Leibnizian space-time by adding a 'standard' of rotation. That is, we choose a rigid reference frame (say, one where there are no inertial effects due to rotation) and declare that it is not rotating. This gives us a collection of preferred rigid reference frames which will include that frame and all other frames that are not rotating with respect to it. As such, in this space-time, it is meaningful to ask questions like:

[4] The idea being that a notion of *relative* rotation would mean that given two extended objects rotating with respect to one another, we could treat one as rotating and the other as non-rotating (or *vice versa*). But, this can't be the case, since the inertial effects of rotation are only experienced by *one* of the objects and so we need to be able to say which one *is* rotating and we can't (at least *prima facie*) do this if we only have a relative notion of rotation.

- 'Is this [extended] object rotating?'. But, it is *not* meaningful to ask 'Is this object moving in a straight line accelerating?'

So, although we now have a notion of *absolute* rotation in order to account for the results of the 'bucket argument', we still can't say whether a particle travelling in a straight-line is accelerating or not. That is, we don't have a notion of *absolute* rectilinear acceleration[5] which is another source of inertial effects. The next space-time attempts to rectify this situation by adding a notion of absolute [rectilinear] acceleration.

Neo-Newtonian (or Galilean) space-time

This is obtained from Maxwellian space-time by singling out a 'privileged' non-rotating rigid reference frame (say, one where there are no inertial effects due to rectilinear acceleration). This gives us a family of preferred frames which we call the *inertial* reference frames and these are the frames whose world-lines are straight (i.e. they are not undergoing rectilinear acceleration) relative to this 'privileged' non-rotating frame. In this space-time, it is meaningful to ask questions like:

- 'Is this particle accelerating?' and this can be asked regardless of whether other particles exist. But, it is *not* meaningful to ask: 'What is the velocity of this particle?', or indeed, 'Is this particle moving?'

So, we now have a notion of absolute [rectilinear] acceleration, and so this space-time is rich enough to describe the situations where absolute accelerations can be detected due to the existence of inertial effects. As such, this seems to be the 'strongest' space-time which is consistent with what we can actually observe about absolute motions. But, a Newtonian wants to have a notions of absolute velocity and position too (even though they have no observable consequences) and the next space-time allows us to have these by adding an absolute frame of reference.[6]

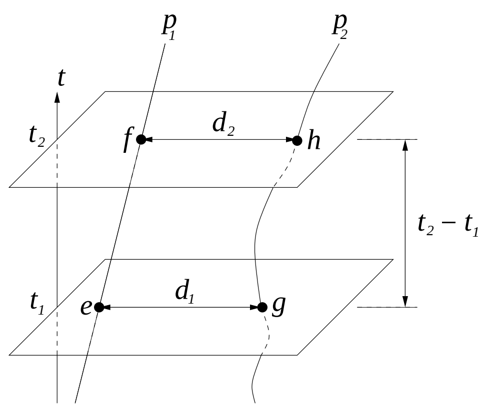

Figure 5: *Neo-Newtonian space-time: In terms of spatio-temporal structure this is just like Leibnizian space-time, but we now specify a special non-rotating rigid reference frame which allows us to define the inertial frames of reference. As such, we say that world-lines which are* straight *relative to such frames correspond to particles that are not accelerating (in an absolute sense). Thus, having fixed such a frame we can say that p_1 is not accelerating, whereas p_2 is.*

Full Newtonian space-time

This is obtained from neo-Newtonian space-time by introducing an absolute frame of reference. That is, by choosing a particular inertial frame (ideally, the one which is at rest relative to absolute space) and declaring that it is at rest. Now, since this absolute frame of reference provides a preferred way of identifying spatial locations through time, it is now meaningful to ask of any particle:

[5] That is, acceleration in a straight line as opposed to acceleration due to rotation.
[6] And, this frame of reference will be the one which we take to be at rest relative to absolute space.

- ‘Is this particle moving?’. And, if it is, we can ask ‘How fast is it moving?’ [relative to absolute space].

As such, this space-time has all of the structure that Newton wanted to utilise in his theory of motion.

Aristotelian space-time (for interest)

This is obtained from full Newtonian space-time by choosing a ‘privileged’ location within the absolute space. In Aristotelian terms, this point corresponds to the ‘centre of the universe’,[7] the point which is needed to define the directions of the ‘natural motions’ of bodies. (For example, bodies composed of ‘air’ tend away from it and bodies composed of ‘earth’ tend towards it.) It is now meaningful to ask of any particle:

- ‘How far is this particle from the centre of the universe?’ and ‘Is this particle tending toward the centre of the universe?’.

As such, we can now draw the Aristotelian distinction between ‘natural’ and ‘forced’ motions and describe, in these terms, why certain material objects follow the paths that they do.

2 The historical background to the Special Theory of Relativity

We now turn to the historical background to Einstein’s *Special* Theory of Relativity. We start by looking at Maxwell’s theory of electromagnetism and its initial interpretation which utilised a ‘luminiferous ether’. In the nineteenth century, it was hoped that this ‘ether’ would provide us with an absolute frame of reference which, in turn, would vindicate Newton’s view of space. And, furthermore, it was hoped that electromagnetic phenomena (as opposed to the mechanical phenomena described by Newtonian mechanics) could be used to detect the elusive uniform absolute motions that are one of the major weaknesses of the absolutist view. This leads us to the Michelson-Morley experiment which was designed to detect such motions by electromagnetic means and the fact that no such motions were detected. Faced with this ‘null result’, two different proposals were offered. As we shall see at the end of this lecture, the first of these, due to Lorentz and Fitzgerald involved postulating a seemingly *ad hoc* ‘length contraction’ hypothesis which implied that the motions in question were still unobservable. In the next lecture, we shall examine the second proposal which is Einstein’s Special Theory of Relativity, and as we all know, this was the proposal that was adopted by physicists.

2.1 Maxwell’s theory of electromagnetism and the luminiferous ether

As we saw at the beginning of the course, Maxwell’s theory of electromagnetism allows us to view light to as a certain kind of electromagnetic wave. And, in turn, these waves were initially thought to travel through a certain medium, called the *luminiferous ether*. This ether, according to the theory, had to have certain properties:

- It had to be diffuse enough to account for the fact that material bodies could travel through it with negligible resistance, and yet strong enough to account for the fact that light travels through it at a very high speed.
- It had to be such that the speed of light only depended on the properties of the ether itself and, as such, the velocity of light is independent of the motion of its sources through the ether.
- It had to be isotropic since the speed of light is the same in all directions.

Consequently, it was claimed that the frame of reference which is at rest relative to the ether could be identified with Newton’s frame of absolute rest. That is, couldn’t we take the ether, an entity posited on *physical* grounds, as the basis for Newton’s absolute space, instead of some metaphysically dubious substance?

[7] Indeed, for Aristotle, this point was the centre of the Earth.

Furthermore, this gives rise to the possibility that we could measure the absolute uniform motion of certain objects. For, even though the absolute velocities corresponding to uniform *mechanical* motions cannot be measured according to Newtonian Mechanics, we should be able to measure the velocity of light in all directions (a non-mechanical *electromagnetic* motion) and hence infer how fast the measuring apparatus is moving with respect to the ether. Obviously, if this could be done, then one of the major objections to the existence of absolute space, i.e. the fact that uniform motions are undetectable, would be rebutted.

2.2 The Michelson-Morley experiment

The Michelson-Morley experiment was designed to do just this: to detect what state of uniform motion was, in fact, the state of rest relative to absolute space. The thinking behind the experiment runs as follows:

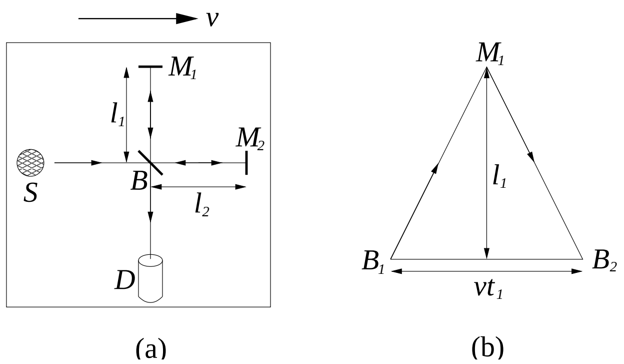

Figure 6: *The Michelson-Morley experiment. (a) A beam of light is sent from the source S to the 'beam-splitter' B which splits it into two beams that are perpendicular to each other. One travels to the mirror M_1 the other to the mirror M_2. After being reflected by the appropriate mirror, both beams return to the 'beam-splitter' B where they are 'recombined' and sent to the detector D. (b) the path of the light beam that travels from B to M_1 and back again.*

The journey from B to M_1 and back: To find the time taken for the light to traverse this distance, we note that the light has to follow the path given in Figure 6 (b). Here, if the time taken for the light to go from the 'beam-splitter' (i.e. B_1) to M_1 and then back to the 'beam-splitter' (B_2) is t_1, then the base of the triangle needed to give the distance B_1M_1 is $vt_1/2$ and so, by Pythagoras' theorem, this distance is

$$\sqrt{l_1^2 + \left(\frac{vt_1}{2}\right)^2}.$$

Then, due to the symmetry of the situation, multiplying this by two gives us the total distance travelled by the light in the time given by t_1. Consequently, since the speed of light is c, we have

$$ct_1 = 2\sqrt{l_1^2 + \left(\frac{vt_1}{2}\right)^2},$$

which, on rearranging, gives

$$t_1 = \frac{2l_1}{\sqrt{c^2 - v^2}}.$$

The journey from B to M_2 and back: To find the time taken for the light to traverse this distance, we break the journey into two parts:

- Let t_{BM_2} be the time it takes the light to go from B to M_2. In this time, the light has to travel a distance $l_2 + vt_{BM_2}$ due to the fact that the mirror is moving away from the light at a speed of v. As such, since the speed of light is c, we have

$$ct_{BM_2} = l_2 + vt_{BM_2} \implies t_{BM_2} = \frac{l_2}{c - v}.$$

- Let t_{M_2B} be the time it takes the light to go from M_2 to B. In this time, the light has to travel a distance $l_2 - vt_{BM_2}$ due to the fact that the 'beam-splitter' is moving towards the light at a speed of v. As such, since the speed of light is c, we have

$$ct_{M_2B} = l_2 - vt_{M_2B} \implies t_{M_2B} = \frac{l_2}{c+v}.$$

Consequently, the time taken for the light to go from B to M_2 and back is

$$t_2 = t_{BM_2} + t_{M_2B} = \frac{l_2}{c-v} + \frac{l_2}{c+v} = \frac{2l_2c}{c^2-v^2}.$$

The time difference: As such, if we set the experiment up so that both arms of the experiment are the same length, say l, then we can see that it takes *different* times for the light to travel these two equal distances. In fact, the time difference in question is given by

$$\Delta t = t_2 - t_1 = \frac{2lc}{c^2-v^2} - \frac{2l}{\sqrt{c^2-v^2}} \approx \frac{lv^2}{c^3},$$

as v, the speed of the Earth through the ether is much smaller than c. Now, given this time difference, although prior to entering the 'beam-splitter' the light was in phase, when it left the 'beam-splitter' after its journey along the arms of the experiment it would be out of phase. As such, interference fringes would be seen at the detector (D in Figure 6).

Rotating the experiment: If the experiment is now rotated through 90°, we can re-run the above calculations to find that the time difference is now given by

$$\Delta t' = \frac{2l}{\sqrt{c^2-v^2}} - \frac{2lc}{c^2-v^2} \approx -\frac{lv^2}{c^3},$$

Thus, there will still be interference fringes, but rotating the experiment in this way will cause these fringes to 'shift' by an amount

$$\delta = (\Delta t - \Delta t')\frac{c}{\lambda} \approx \frac{2lv^2}{\lambda c^2},$$

where λ is the wavelength of the light being used. As such, performing this experiment and rotating it, Michelson and Morley could measure this 'shift' and hence determine v, the velocity of the Earth relative to the ether.

But, on performing the experiment, Michelson and Morley found *no 'shift'*, i.e. there was *no detectable difference* in the time taken for the light to traverse the two perpendicular distances when the experiment was rotated. As such, it would seem that the velocity of light is independent of our state of motion relative to the ether. Clearly, this challenges one of our fundamental intuitions about motion, namely that when we 'chase after' a moving thing, say a material body or a sound wave, it will be moving more slowly with respect to us than someone who chose not to 'join the chase'. Thus, this 'null result', i.e. the non-appearance of an expected effect, requires an explanation.

2.3 The Lorentz-Fitzgerald hypothesis: length contraction

To explain the 'null result' in question, the physicists Lorentz and Fitzgerald proposed the following hypothesis:

> When in motion with respect to the ether, the apparatus *shrinks* in length by some appropriate factor.

In particular, it was claimed that the length of the arm in the direction of motion relative to the ether should shrink by a factor of $\sqrt{1-v^2/c^2}$, i.e. l_2 in our earlier calculation should 'really' be $l_2\sqrt{1-v^2/c^2}$. As such, the time t_2 which was

$$t_2 = \frac{2c}{c^2-v_2}l_2 \text{ becomes } t_2 = \frac{2c}{c^2-v_2}l_2\sqrt{1-v^2/c^2} = \frac{2l_2}{\sqrt{c^2-v_2}},$$

which, on the assumption that $l = l_1 = l_2$, means that we now have $\Delta t = \Delta t'$.[8] As such, in the presence of the 'length contraction' hypothesis, we would expect the 'shift', as measured by the difference between Δt and $\Delta t'$ to be zero, as observed!

However, even though the 'length contraction' hypothesis gives us a way of 'correcting' our analysis so that it agrees with the results of the experiment, it does so at some cost. To see why, notice that in the presence of the 'length contraction' hypothesis:

- the relative velocity of two things is not simply the [vector] sum of their individual velocities.[9] That is, in such cases, we lose the intuitively obvious 'classical' law for calculating relative velocities.
- the observed velocity of light is independent of our state of motion relative to the ether. So, the introduction of this 'new physical effect' allows us to explain the null result, but unfortunately, it also seems to rule out the possibility of measuring absolute velocities by using such electromagnetic phenomena.

Thus, the 'length contraction' hypothesis 'saves the phenomena', but is the cost acceptable?

2.4 How Einstein's programme superseded Lorentz's and the *ad hoc* nature of the Lorentz-Fitzgerald hypothesis

There is a sense in which Lorentz and Fitzgerald's 'length contraction' hypothesis is clearly *ad hoc*. That is, no-one suspected that such an effect was necessary prior to the experiment and it seems to be introduced, by and large, to account for the experimental result. As such, this seems to be a classic case of trying to 'save' a theory by introducing an *ad hoc* assumption. However, some philosophers (most notably, Grünbaum[10] and Zahar[11]) have argued that this is not so. This is not a debate that I wish to get into here and so we shall turn, instead, to the way in which Einstein accounted for the null result.

James Ward (e-mail: `j.m.ward@lse.ac.uk`)

[8]In this special case, notice that Δt and $\Delta t'$ are both zero, i.e. we wouldn't actually get any interference fringes! But, in the general case, where $l_1 \neq l_2$, we have $\Delta t = \Delta t' \neq 0$, i.e. we get fringes but no 'shift'.

[9]At least, if one of them is a beam of light.

[10]See A. Grünbaum, *The falsifiability of the Lorentz-Fitzgerald contraction hypothesis*, British Journal for the Philosophy of Science, vol. 10 (1959).

[11]See E. Zahar, *Why did Einstein's programme supersede Lorentz's?*, British Journal for the Philosophy of Science, vol. 24 (1973).