

Understanding Relativity

or

How to do Effective Thought Experiments

Ian M. Sefton
School of Physics, The University of Sydney
I.Sefton@physics.usyd.edu.au

Introduction

After some years absence, the topic of special relativity has reappeared in the NSW HSC Physics syllabus as part of the wonderfully broad course module, *Space*. That is appropriate because next year, 2005, we will be celebrating the World Year of Physics, which coincides with the centenary of Albert Einstein's great original papers on three quite different topics (Brownian motion, the photoelectric effect and special relativity). The topic which we explore here is the special theory of relativity which grew from the seed of James Clerk Maxwell's electromagnetic theory (which is also well-represented in the Stage 6 syllabus).

Part of the fascination of relativity lies in its contradiction of our deeply ingrained intuitions about the nature of time and space. Some people claim that this contradiction makes the topic too difficult for ordinary mortals to understand, but that's an over-reaction; all you need is a little knowledge of physics and the patience to work through a logical argument. When you understand the logic you can decide for yourself whether to trust your intuition or your reasoning. In this article I will discuss several aspects of special relativity in more detail than that found in most school-level texts. By working carefully through these arguments you should be able to improve your conceptual understanding of the subject.

Thought experiments and measurements

Following a tradition established by Einstein himself, many writers on relativity have employed a technique for conceptual reasoning called the thought-experiment (Gedankenexperiment in German). The thought-experiments of special relativity are performed in an idealised abstract world with simple apparatus operated by imaginary observers who work in different frames of reference. Typically, the different frames have very high relative speeds, so an appropriate modern example would use a spaceship and an Earth-bound lab as suitable frames, but the syllabus (following Einstein) specifies trains instead of space-ships. The basic instruments used for thought-measurements are clocks and rulers and a key requirement is that each observer uses apparatus which is at rest in her/his own frame. Imagine all clocks and rulers as being glued to their own frames. Importantly, we also assume that all clocks have been calibrated against each other and all rulers are identical.

First we need to get an initial grasp of three important concepts: *frame of reference*, *event* and *observer*. You can imagine each frame of reference (frame for short) as being something with a fixed system of co-ordinate axes. All the objects in a given frame remain in fixed positions relative to each other. Different frames can be moving relative to one another; indeed most of our reasoning will involve analysis of the same set of events as they are observed in two different frames. An event is something that happens at a specified location, idealised as a point, at a specific instant of time in some frame. In another frame the same event will have a different location and time. Observers observe events. Every observer stays at one location in one frame only. Therefore, in order to make good observations of events which happen at different places, we need many observers in every frame. (Books which describe examples using only one observer in each frame are usually wrong.) One of an observer's jobs is to record the time at which an event occurs. For that reason, I will often represent an observer by her clock.

It is also important to distinguish between a time reading on a clock and a time interval, which is the difference between the readings from two clocks in the same frame. Quite often the

measurement of a time interval will require two clocks (with two observers) at different locations in the same frame. One way of emphasising this important difference is to use the notation t, t' for times and $\Delta t, \Delta t'$ for time intervals. Unfortunately the syllabus document and the texts that follow it do not follow that convention, so you will need to sort things out from contexts. (The time dilation formula on page 44 of the syllabus refers to time intervals.)

Similarly, an observer can record the location of an event on a ruler, but to measure a length we need two observers at different places on the same ruler.

One term that can cause trouble is “inertial frame”. A simple definition is that an inertial frame is one in which Newton’s laws of motion are valid. That can be a worry because the third law proposes that every force is one member of an action-reaction pair, an interaction between two bodies, and it is implied that the two forces are always exact opposites; if an action changes then its partner, the reaction, will also change simultaneously even if the bodies are far apart like the Earth and the Moon. That newtonian idea of instantaneous action-at-a-distance does not fit into relativity, so it is best to forget about the ancient version of Newton’s third law when we talk relativity. It is OK to say that a frame of reference in which Newton’s first law is valid is an inertial frame. The crucial property of an inertial frame in special relativity is that, once you have found one frame that qualifies as inertial, any other frame moving relative to that one with constant velocity is also inertial. It is sometimes written that an accelerated frame of reference can’t be an inertial frame, but physicists writing in the context of general relativity will say that a system in free fall, such as an orbiting satellite, can also be an inertial frame. That view is a reflection of the idea that there is no practical difference between having a locally uniform gravitational field and being in an accelerated frame.

Misleading terminology

Some text-book discussions of relativity say that an observer “sees” or “perceives” some value for a physical quantity, but those terms are misleading. Simple relativity theory does not predict what one sees with the eyes, nor does physics deal with perceptions or mental constructs. For example, an observer does not see light travelling along a path but can observe only the sending and arrival of light pulses. A better way is to talk about what is measured or observed directly using instruments. In reading and understanding the text books, therefore, it is a good idea to substitute “measure” for “see”, “perceive”, “observe”, etc. whenever those words occur. If making those substitutions does not make sense then you should suspect that the text has left something out or has got something wrong.

If you do want to talk about seeing then it is important to realise that what one actually sees in many cases is influenced by two things: how long it takes light to get from the source to the eye and the relativistic effects of motion on measurements of distances and times. For example if you were to observe a very long spaceship that is moving away from you, light from the front of the ship would take longer to reach you than light from the back, simply because the front is further away, and that will affect how the ship looks to you. But that is not a relativistic effect. The basic theory of special relativity is concerned about things that are measured, such as the length of the moving spaceship or the rate of a ticking clock carried on board the ship.

Similarly, pictures showing waves and rays of light as “seen” by an observer are not to be taken literally. You can’t literally see waves and rays. Such diagrams are conceptual models, constructed by the mind’s eye to represent explanations of the measurements. By the way, the same goes for pictures of wave fronts and rays in optics books; those waves and rays are not what we see or measure directly.

The basic principles of special relativity

The whole of Einstein's special theory of relativity theory is based on classical mechanics, classical electromagnetism and two new propositions.

- The laws of physics are the same in all inertial frames.

- The speed of light in vacuum is a universal constant (c) and is independent of the motion of its source.

The first statement is usually called the *principle of relativity* and was more or less understood, at least in the domain of mechanics, before Einstein. Einstein extended it by taking in the laws of electromagnetism. The audacious, exciting, contribution by Einstein is the new conception of the nature of light contained in the second proposition. That universal speed of light is part of the structure of the universe, not to be mucked with by any experiment.

A hundred years after Einstein we have become rather blasé about the universality of the speed of light; it has become so well established in the belief-system of physicists that it is now almost a tautology. It has been decreed that in the SI system c has a fixed value, specified with a precision of 9 significant figures. The meter is now defined in terms of the second and the value of c . So nowadays any direct “measurement” of c automatically involves SI units of both time and distance that are guaranteed to give the decreed value.

Time dilation and the light-clock

A typical example of a thought-experiment is the analysis of the “light-clock”, discussed in many texts, which can be used to derive the “time dilation” formula. The basic idea of a light-clock is to use the distance travelled by a pulse of light and the known speed of light to mark out intervals of time. The left-hand diagram of figure 1 models what goes on. A flash of light is sent out for a distance L and is reflected from a fixed mirror back to its starting point. The light-clock “ticks” by sending out a new light pulse every time that one comes back. A single conventional clock in the thought-laboratory of frame S , located beside the sending and receiving apparatus (which is assumed to be small compared with the distance to the mirror), is used to measure the time interval (the period) of a tick. In the rest frame, S , of the apparatus it is easy to see that the period must be $2L/c$.

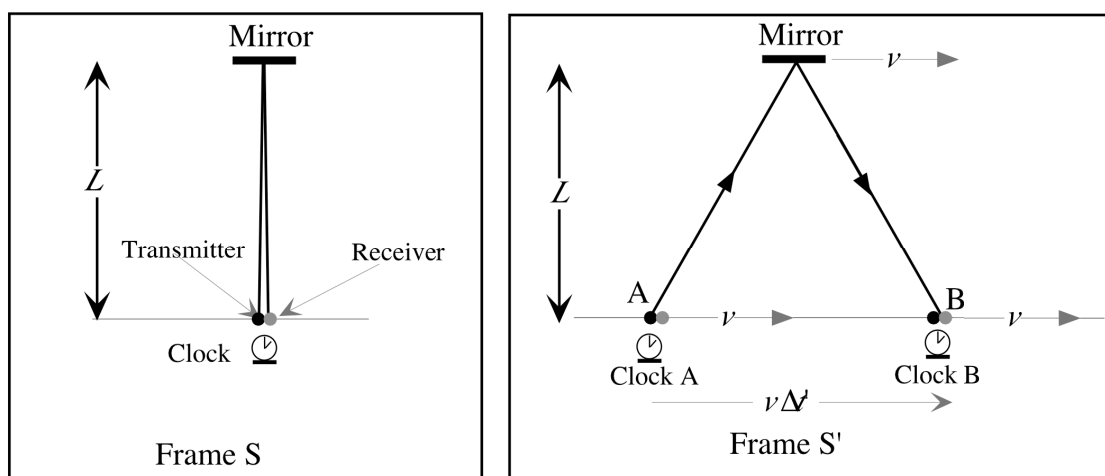


Figure 1. Measurements and analysis for the light-clock made in different frames

Left: No relative motion. Right: light-clock moving at speed v . Two ordinary clocks are required in frame S' since they must both be at rest in S' . Analysis of these diagrams leads to the time dilation formula.

In the second part of the thought-experiment the light-clock moves rapidly sideways with a speed v measured in our frame, S' . The right-hand diagram shows how observers in frame S' model and measure the operation of the moving light-clock. Because all instruments used to take measurements in our frame S' must be at rest in that frame we now need two conventional clocks in the thought-laboratory in order to measure the period of the light-clock’s tick. We need one clock (A) at rest in our frame to record the time that a light pulse is sent out and, since, the light-clock moves on, a second clock (B) at another location to pick up the returning light pulse.

Furthermore we need to know that those two clocks are synchronised; they have to agree on the time. (We will come back to the problem of how to synchronise those clocks.)

The analysis depends on the proposition that the speed of light is the same for everybody. An observer travelling with the light-clock (frame S) who wants to compare the period of the light-clock's tick with a conventional clock needs only one clock, because the situation is exactly the same as in the left-hand part of figure 1. The traveller's value for the tick period is

$$\Delta t = t_0 = \frac{2L}{c}.$$

(Δt is a time interval measured in frame S travelling with the clock; t_0 is the notation for Δt in the syllabus document.)

For us, in the thought-laboratory frame S' with our two clocks, the result is quite different. It is immediately obvious from the diagram that the measured period of a tick, $\Delta t'$, must be longer than Δt . It is also fairly easy to work out the so-called time dilation formula for the tick period measured with the two stationary clocks:

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \text{or in syllabus notation,} \quad t_v = \frac{t_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

These two time intervals are clearly different, and the thought experiment forces us to the conclusion that there is no such thing as universal time. Time depends on the frame of reference in which you measure it.

There is a symmetry in this argument: if the observers travelling with the original light-clock in frame S make similar measurements on a light-clock that is glued to our frame S' they get the same answer: they measure our clocks as going slow. We use our stationary clocks to measure their clocks as running slow and they use clocks at rest in their frame to measure our clocks ticking slow. That result implies that one inertial frame is as good as any other; both sets of measurements are equally valid.

Did you spot anything fishy in this thought-experiment? Did the arrangement with two clocks fortuitously placed in exactly the right positions of frame S look like a fudge? Clearly, to place the two clocks correctly we need to know what the result of the thought-experiment is going to be. That's no big deal – it is OK to predict the result of a thought-experiment. An alternative approach, which can be used in thought-experiments when we don't know the answer from theory, would be to have a whole array of clocks strung out along the path of the moving apparatus. We could then use the two clocks which happen to be in the right places when the light-clock's signal goes out and comes back again – rather expensive in the real world, but it is a thought-experiment after all. But there is another issue here: not only do we need two synchronised clocks but we need two thought-observers in the right places to read each of clocks. Those "observers" don't have to be human; we could equip each clock with something like a video recorder which will keep a timed record of everything that happens. Better still, we might save thought-money by replacing each clock with a simple detector that feeds a signal back to a central processor. Although it takes time for those signals to reach the processor, the processor can correct for those time delays using the known signal speed and the location of each detector. The point of this argument is that what is meant by a clock in a thought-experiment can be flexible as long as we have some kind of stationary detector at all relevant locations. We can still represent the whole time-measuring apparatus by an array of stationary recording-clocks and it remains true that we need at least two of those clocks to do the light-clock experiment described above.

Exercise: Synchronising a set of clocks

A useful intellectual exercise is to invent a method for synchronising all the clocks needed for the thought experiments. One idea is to do as they do in adventure stories: gather all the clocks at a central thought-laboratory and synchronise them there before dispersing them to their final

locations. But somebody might object that the act of moving the clocks to their allotted locations could alter their readings; after all we are proposing thought-experiments to test what happens to moving clocks. A better procedure would be to put all the clocks in place and measure their positions before they are synchronised. The time intervals required for light to travel between them can then be calculated. Then, at a predetermined time on some master clock, we send out one light signal (a spherical wave) from a central location to all clocks. Each clock can then be set to the appropriate time when it receives the signal. I leave the rest of the details for the reader to work out.

When you are happy that it is possible to synchronise clocks, consider this: a set of clocks that is synchronised in one frame of reference can be out of synch in another frame. That's our next problem.

Simultaneity

An important result from the special theory is that events that are recorded as being simultaneous in one frame of reference will not be simultaneous when measured in another frame that is moving relative to the first one. This does not mean that remote events are never simultaneous. For example, here is another thought experiment. Suppose that we see a light signal from spaceship *Archimedes* that we already know to be at a distance of 10 light minutes - that's event 1. 10 minutes after event 1 comes event 2: we get another signal from another spaceship *Bernoulli* which we know to be 20 light minutes away. Events 1 and 2 are not simultaneous. Now consider two different events that we did not see: event A is the light pulse leaving *Archimedes* and event B is the light leaving *Bernoulli*. Knowing the speed of light (1 light minute per minute) we can work out that events A and B were simultaneous. The reasoning in this example has nothing specially to do with relativity, the surprise is yet to come.

The relativity of simultaneity is often discussed using a thought-experiment that involves the simultaneous launching of two light pulses from the middle of a long train carriage towards the ends of the carriage (I would prefer to use a space-ship). Andriessen et al (2001) discuss a train with an operator who uses the light pulses to open doors at the two ends of the train. I will discuss the same example because I think that their description is flawed. They write about what the operator in the train “sees” compared with what a single outside observer “sees” but, as in the case of the light-clock, measuring rather than seeing is what matters.

In the train's frame it is true that the operator will actually see both doors open at the same time but that seeing takes place some time after the doors have opened. The fact that the distances from the operator to each of the doors are equal ensures that light signals both ways have the same travel times. But mere “seeing” is not good enough for the observer outside the train; measurements are required.

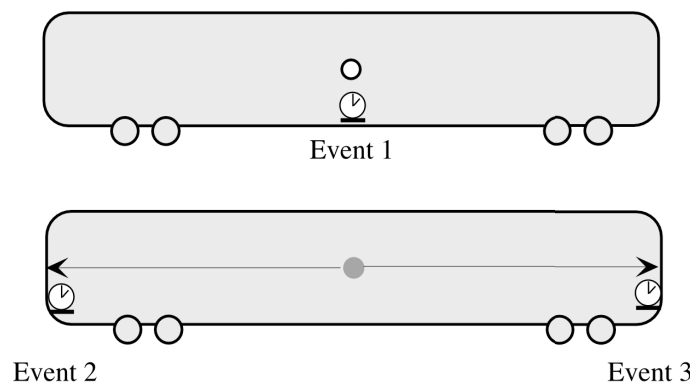


Figure 2. Simultaneity experiment measured inside the train

Event 1 is light leaving the lamp in the middle of the train. Events 2 and 3 are light reaching the ends of the train.

Consider first what happens in the frame of reference of the train (figure 2). We need three synchronised clocks (and three observers to read them). Event 1, the emission of the light flash, is recorded by clock 1 in the middle of the train. Events 2 and 3, the arrival of the light at the rear and front ends of the train are recorded by two more clocks placed at the ends. It is easy to see that in the train's frame, the times for events 2 and 3 must be the same; they are simultaneous.

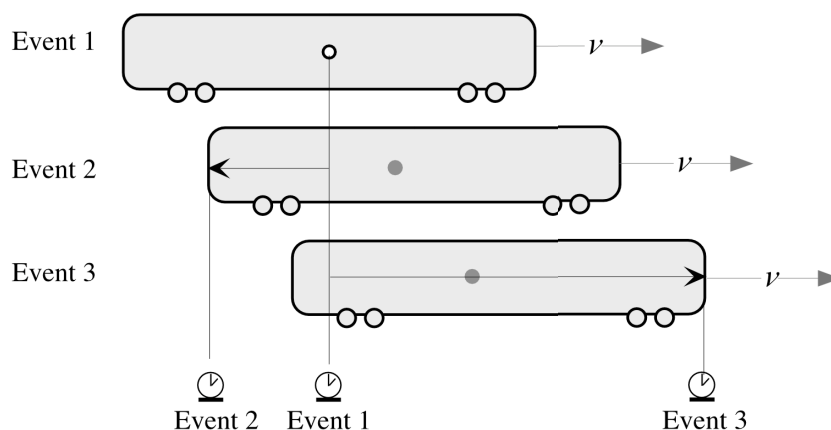


Figure 3. Simultaneity experiment measured from outside the train

Event 1 is light leaving the lamp in the middle of the train. Event 2 is light reaching the back end (short arrow). Event 3 is light reaching the front end (long arrow).

Now consider measurements of the same three events made by three observers with three clocks located outside the moving train. Figure 3 shows the positions of the train for each of the three events. The left-pointing arrow in the second picture shows how far the light has to go to get from its starting point to the back of the train and the right-pointing arrow in the third picture shows the distance travelled by the light to the front of the train. Remember that the speed of light is still c . It should be clear from the diagrams that the light gets to the back of the train first, because the train is moving towards the light flash. The arrival of the light at the two ends is not simultaneous.

It should now be clear why we don't reason about what a single observer "sees" in this example – that is much more complicated. It is, of course, possible to work out or measure when a single observer receives light from the opening doors at the two ends of the train if we know where the ends of the train are at the times of events 2 and 3 (using a big ruler alongside the railway track). We could then work out how long it takes light to get from the ends of the train to the observer's eyes. But why bother? We have already established the lack of simultaneity.

It is also worth noting that if two events occur simultaneously *at the same place*, as measured by one observer, then all observers in other frames will agree that the events were simultaneous.

The relativity of length

The term “length contraction” appears in the syllabus and many texts. What does it mean? Does it imply that a fast-moving object really shrinks? Well no, not exactly. Some texts will tell you that an object only appears to shrink or that the observer “perceives” or “sees” a contraction, but both of those explanations miss the point, as we shall see.

First, where did the term “length contraction” come from? It probably derives from a theory invented, before Einstein, by H. A. Lorentz, which was an attempt to explain things like the negative result of the Michelson-Morley experiment. That was a real experiment designed to detect Earth's motion through the aether that was supposed to be the carrier of Maxwell's electromagnetic waves. The basic idea was that when an object gets dragged through the aether the object shrinks along its direction of motion, an intuitively appealing idea. Lorentz even worked out the formula for the contraction that is quoted in the syllabus and turns out to be the same as the result of Einstein's theory. However, Lorentz's theory was soon discarded in favour of Einstein's explanation.

The key to understanding the relativistic explanation is to ask how, exactly, one might measure the length of a moving object. We do another thought experiment with the same rules about stationary clocks and rulers. Let's adopt Einstein's example of a hyper-express train and see how to measure its speed. We need a long ruler laid out beside the track and, not one clock, but a set of stationary clocks spread out alongside the ruler. See figure 4.

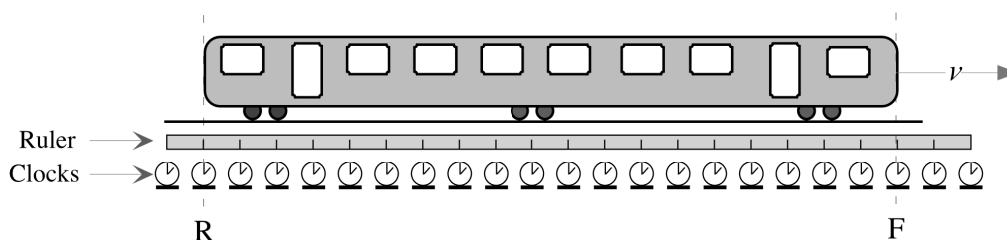


Figure 4. Measuring the length of a moving train

Clocks R and F are required in order to synchronise the measurements of position on the ruler.

All the clocks have been synchronised so that they all show exactly the same time and we know the positions of all the clocks along the ruler. The measurement of length consists of recording the simultaneous positions of the two ends of the train. At the instant depicted, the front of the train is at the position of clock F and the rear end is at clock R (that's two events). The measured length of the train is then equal to the distance between clocks F and R. This thought-experiment also requires more than one “observer” because we need a gremlin or some other helper at each clock to do the actual “seeing” of the end of the train and the reading of the clock. The instructions might be to report in if you see either end of the train at your location at precisely 0 nanoseconds after 1 o'clock. The observers can report later but we can work out, by knowing who reported in, where the ends of the train were at the chosen time.

It may seem odd to have to use clocks to measure distance but their role is crucial: it is to make sure that we measure the positions of the ends of the train simultaneously in our frame.

The point of this discussion is that Einstein's theory predicts that the measured length of the moving train will be less than that for the stationary train. The two measurements are related through the formula quoted in the syllabus:

$$l_v = l_0 \sqrt{1 - \frac{v^2}{c^2}}$$

I think that understanding the argument above is the key to answering the question of whether a moving object “really” contracts. It depends on what you mean by reality. I take the view that the only kind of reality that science deals with is described by experiments, observations and measurement. In that scheme, yes, it does shrink. In another observer’s reality the amount of shrinking may be different. And for observers on the moving train (who use clocks and rulers fixed to the train) the reality is that the train doesn’t shrink. But that discussion is starting to look like philosophy. I think that students will have a better chance of understanding if we refrain from talking about “length contraction”, “seeing”, “perceiving” etc. Instead we should talk about measurements in real experiments as well as in thought experiments.

The twins paradox

Although the twins paradox is not formally required by the syllabus, it is a nice thought-experiment problem and it is included in at least one text (Andriessan et al, 2001) used for the HSC course. A paradox consists of a set of statements which appear to be inconsistent, but which, on closer examination, turn out to be compatible. (So talking about an “apparent” paradox, as some people do, is being redundant.)

The story involves twin scientists, Alice and Bob, and this time we do need a space-ship rather than a train. Alice goes away on a long journey through space and then comes back again, while Bob stays at home keeping an eye on his clock. To keep the physics simple, Alice’s spaceship quickly accelerates and then cruises for most of the outward journey at a speed relative to Earth (Bob’s frame), which is a substantial fraction of the speed of light. She then turns the ship around and travels back to Earth at the same speed. The paradox arises from the times recorded for the outward and return journeys recorded by the Alice’s clock(s) and a clock kept at home by Bob. By applying the time dilation formula to Alice’s two constant-speed journeys we can deduce that the total time measured by Alice’s clocks is less than that recorded on Bob’s clock. (Extending the argument to Alice’s biological clock we can see she has not aged as much as Bob.) But according to relativity, one frame is as good as another, so why doesn’t Alice find that Bob is younger?

A common mistake in resolving the twins paradox is to look for the cause of the Alice’s youthfulness in the accelerations that she and her clocks undergo at the beginning, middle and end of the great journey. But the real reason for the different ages has already been discussed: it is directly attributable to the different time measurements made in **three** different frames of reference. Alice occupies two inertial frames while Bob lives in one frame.

To see that the regular time dilation effect rather than Alice’s accelerations is the correct explanation, consider a variation on the thought-experiment. Alice travels at the same speed as previously on the two legs of the journey but goes twice as far each way. The time dilation effect will be exactly doubled. But Alice can accomplish the extended journey using exactly the same accelerating manoeuvres at the starting, turning around and stopping stages. So whatever effect the accelerations may have on Alice’s time (and there is an effect predicted by general relativity) we can make that effect negligible by extending the total journey. It is possible to argue this case in much finer detail, working out the details of what Bob’s observers would measure on Alice’s clocks and what Alice’s observers would measure on Bob’s clocks, but the conclusion stands: Alice is younger than Bob. To follow up on those details borrow a good book such as Sartori (1996).

Finally, it has to be said that there are some people who think that the laws and principles of physics don’t apply to biological systems. (That’s not the same as saying that physics can’t yet “explain” all of biology, which is clearly true.) In this argument I think that whatever physics has to say about the nature of time and clocks must also be applicable to biological clocks.

Acknowledgements

I thank Ian Cooper, Bob Hewitt, Brian McInnes and Ann Sefton for helpful comments and suggestions.

Feedback

I would like to know whether you found this article useful. Please email any comments or questions to I.Sefton@physics.usyd.edu.au.

Bibliography

Andriessen, M., Pentland, P., Gaut, R., MacKay, B., McCallum, Y. (2001). *Physics 2 HSC Course*. Milton, Qld: John Wiley.

Board of Studies, New South Wales (2002). *Physics Stage 6 Syllabus, Amended October 2002*.

Sartori, Leo (1996). *Understanding Relativity, A Simplified Approach to Einstein's Theories*. Berkeley: University of California Press.

APPENDIX: GLOSSARY

Since the explanations in this glossary are all tied to the idea of making measurements and doing thought-experiments in special relativity, the meanings of some terms are much narrower than usual.

Clock: an instrument used to measure time. Unlike stopwatches, all the clocks that are used in one frame need to be synchronised and at rest in the frame. It is generally assumed that all clocks are identical.

Event: an occurrence at some point in space at one instant of time. It can be described by one set of co-ordinates (x, y, z, t) in a specified frame.

Frame of reference: a set of objects which remain at rest relative to one another, to which we can attach a co-ordinate system and a timing system consisting of a set of clocks.

Inertial frame: a frame in which Newton's first law is valid. Given any inertial frame S , then any other frame S' moving relative to S with constant velocity is also inertial. (It is *not* true that absence of acceleration is required - a lift in free fall is as good an inertial frame as the surface of the Earth.)

Length contraction: the factoid that the measured length of a moving object is less than the measured length of an identical object at rest.

Observer: a thought-person who measures events. Every observer needs a clock. Observers and their clocks are always at rest in their own frame.

Principle of relativity: the assertion that all the laws of physics are identical in all inertial frames.

Proper time: the time interval between two events that occur at the same place.

Rest frame: the frame in which an object of interest is at rest. There is no such thing as an absolute rest frame. (It is **not** "the frame in which a measured event occurs"; a single event can be observed or measured from many frames.)

Ruler: a thing for measuring positions $(x, y, z,)$ in some frame of reference.

Simultaneous events: events which occur at the same time in a specified frame. Events which are simultaneous in one frame may not be so in another frame.

Thought-experiment: an experiment conducted in the mind and with mathematics in order to illuminate and study the principles of physics.

Time: what a clock measures (such as 3 milliseconds after half past one).

Time interval: the difference between two times measured in the same frame

Time dilation: the conclusion that the period of a moving clock is measured to be slower than the period of an identical stationary clock.

Velocity of light: a sloppy way of saying speed of light. Unless the context indicates otherwise, it is assumed to be the speed of light in vacuum. It is a universal constant.