A Few Off-the-Beaten-Track Observations and Challenges

in

Economics, Physics, Computer Science and Mathematics

by

Peter Schorer

peteschorer@gmail.com

Apr. 14, 2022

This chapter from the forthcoming second edition of *Shaving With Occam's Razor* contains thoughts, questions, and projects that have occurred to me in recent years. The answers to some of the questions may be well-known.

Note: I am deeply indebted to Ed Boyda, a young physicist, for his extensive and insightful responses, via email, to many of the items in this chapter. His responses are identified by his initials, "E. B."

I will welcome hearing from other readers with relevant information.

Economics

Note: Further observations on economics will be found in "Politics and Economics" in John Franklin's collection of essays, *Thoughts and Visions*, on thoughtsandvisions.com.

Mass Production

Why is an assembly line, with each worker performing the same task over and over, more efficient for producing large quantities of the same product? Assuming that a worker who made the product entirely himself were equally skilled at performing all stages of the work, is the advantage of the assembly line that there is, for all practical purposes, no cost in time of converting from one stage of the work to another? Give answer in mathematical terms.

The Two-Investment Problem

A person has two investments, one earning interest at a lower rate than the other. Interest is compounded every interest period and immediately thereafter an amount is withdrawn from one or both investments. The amount is always the same. Suppose the amount is greater than the interest produced each period by the lower-interest investment. Should the investor (a) make up the difference by withdrawing, each period, from the interest yielded by the higher-interest investment, or (b) should he draw down the lower-interest investment first, and only then start withdrawing from the interest yielded by the higher-interest investment? Many people argue that (b) is the correct answer.

Assume inflation and taxes and bank fees are all zero. Assume, also, that no money can be deposited into the higher-interest investment apart from interest earned by it during a given period. (Otherwise, the problem becomes trivial.) In other words, the higher-interest investment is like an IRA except that not even a $2,000 annual deposit is allowed. Money *can* be deposited into the lower-interest investment provided it comes from the higher-interest investment.

Assume that the initial investment amounts, the interest rates, and the withdrawal amount can take any value greater than or equal to 0, subject to the indicated constraints. (Of course, if the withdrawal amount is too large, the solution becomes trivial.) This means that the set of all possible cases is continuous, i.e., for any case under discussion, we can specify a case (many cases!) arbitrarily close to, but different from, it in one or more of the parameters.

An argument that (a) is the best answer *under certain circumstances* is the following: Suppose that the withdrawal amount is just "slightly" larger than the interest earned each period by the lower-interest investment. Furthermore, suppose that, by transferring a very small amount *once* from the larger-interest investment to the smaller, the investor can make the withdrawal amount *exactly equal* the interest earned each period by the lower-interest investment. Then, after the transfer: (1) the investor can continue to withdraw that amount forever from the lower-interest investment, and (2) the higher-interest investment will grow forever, and, furthermore, it will grow only "slightly" less fast than it would have without the transfer. But this is clearly not the same as (b), above.

Unfortunately, this argument doesn't show that this is the best strategy, meaning, that it will result in the largest total investment in the long run — or, more precisely, that from some point on, the total investment will grow more rapidly than will the total investment under all other strategies for the same initial parameters. It only suggests that the strategy permits the total investment to grow arbitrarily large with time.

A Good Way to Distribute Money into Investments of Varying Risks and Returns

Investment advisors tell us that the first rule of responsible investing is to diversify our investments, and, in particular, never to invest a large proportion of our money in investments that are risky. But at least in the popular press, and on TV programs devoted to financial management, we never hear anything more precise than that. The question is, can we in fact be more precise about distributing our money among investments of varying risks and returns?

Suppose we have a certain amount M of money we want to invest, and suppose we have $n \geq 1$ different choices for investments, each with a certain risk and a certain rate of return. We will denote our investments inv_1, inv_2, ..., inv_n, and we will denote the risk associated with inv_i, $1 \leq i \leq n$, by $risk(inv_i)$, and the rate of return of inv_i by $return(inv_i)$.

The question immediately arises, how will we assign a numerical value for risk to each of our investments? We normally assume that risk "is proportional" to return, but we are seldom if ever more precise than that. Does risk increase linearly with return, so that, e.g., twice the return implies twice the risk? What does "twice the risk" mean? That we are twice as likely to lose all of our investment, or twice as likely to suffer any loss at all, or ...? Or is the relationship non-linear, so that, e.g., risk goes up as, say, the square of the return? I know of no generally-agreed-upon answer to these questions, so, initially at least, let us assume that risk increases linearly with return, and that the proportionality constant is 1. We will arbitrarily assign a risk value of 1 to our safest investment, then, to compute the risk value for any other investment, we will apply the formula,

$$risk(inv_i) = return(inv_i)/return(inv_1)$$

Thus, if our safest investment returns 6%, (hence, by assumption has a risk of 1), then if we have an investment that returns 12%, we will assign a risk of 2 to it.

We now must decide on a total return that we want. Clearly, it must be ≥ our lowest return, and ≤ our largest return. But, as far as I know, there is no generally-agreed-upon formula that, given a set of risk/return pairs, will tell us which such pair is "best" for us. So we must make the decision ourselves, based on our needs and the risks we are prepared to run. The program outlined below will examine all possible combinations of investments and amounts of money for each investment, and tell us all the combinations that give us the return we have specified, along with the risks associated with each. We can then choose the combination or combinations that have the lowest risk.

Before I describe the very simple program, I would like to mention an idea that occurred to me during the course of writing this sub-section. It is this: that instead of regarding a high rate of return as a way to make a lot of money on a given principal, instead, *we regard a high rate of return as a way to invest less money to achieve a given absolute amount of dollars in return*. If we begin by deciding on the overall rate of return we want from our investments, then this idea suggests that a program such as we are about to describe will, in fact, enable us to achieve the lowest possible risk for the rate of return we have specified, because the program will be able to assign the smallest proportions of our money to the highest risk investments, these small proportions producing as many dollars as large proportions of low risk investments.

So now to our program, which is written in pseudo-Pascal. We give the top level only.

desired-total-return is the return we have decided we want.

num-segments is the number of segments into which we choose to break down our total amount *M* of money to be invested. The larger that *num-segments* is, the finer the discrimination between investment possibilities.

a *partition* of *M* is a grouping of connected segments of *M*. There are $2^{num\text{-}segments - 1}$ possible such groupings because there are *num-segments* - 1 dividing lines between two successive segments. Assigning 0 to such a dividing line if the two successive segments are to be in the same group of segments, and 1 if the two are to be in separate segments, the total number of possible groupings is seen to be $2^{num\text{-}segments - 1}$.

begin program

for each possible *partition* of *M* do
    for each possible assignment of investments to each grouping in the partition do
        begin
            Compute *total-return* for that assignment of investments;
            Compute *total-risk* for that assignment of investments;
            Save *total-return* and *total-risk* in next location in *return-risk table* along with
                assignment of investments and partition;
        end

Go through *return-risk table* and print out the rows that have the *desired-total-return*.

end program

The program at its conclusion gives us all the investment combinations that give us the total return that we desire. We can then choose the one, or several, among these, that have the lowest risk. Observe from the program that it is possible that one or more of these *desired-total-return* investment combinations might not include all *n* of our investments.

As far as I know, no such program is readily available on the commercial market, and yet it enables us to sharpen our investment decisions considerably.

Physics — Special Relativity

Note: probably the most interesting sub-section in this section is the one that questions a basic precept of Special Relativity: "Conjecture on Detecting Whether or Not an Inertial Frame is Moving." on page 12

Plausibility Arguments for Some Basic Facts of Special Relativity

A neighbor of mine said he would be willing to learn something about Einstein's Special Theory of Relativity as long as it could be done with no math, no drawings, and no required reading or study on his part. It was a challenge I couldn't resist. The following is a written version of my attempt (which, although I had to use a few numbers, he tentatively felt was successful).

The Two Basic Assumptions of Special Relativity

Einstein began with several assumptions, among which are: (1) that the speed of light — 186,000 miles per second, or 300,000,000 meters per second — is a constant throughout the universe, and (2) that nothing can travel faster than the speed of light.

Evidence that (1) was in fact true had been obtained in the 1880s through the Michelson-Morley experiment, "which Einstein later said repeatedly had no direct effect on him, and of which he may not even have been aware"[1]. To say that the speed of light is constant is to say that, no matter where you measure the speed of light — whether, e.g., you are at a stationary point, and the light beam is traveling in a rocket ship moving at a fixed velocity, or whether, e.g., you are traveling in a rocket ship moving at a fixed velocity and the light beam is traveling from a stationary point, or whether you and the light source are both stationary relative to each other, or whether you and the light source are moving — you must always get the same value.

The Three Basic Facts Established by Special Relativity

Working from assumptions (1) and (2), Einstein came up with three conclusions, namely, that as an object — say, a spaceship — approaches the speed of light:

(A) time in the object slows down relative to a stationary observer;
(B) the length of the object decreases, relative to a stationary observer, in the direction it is traveling; and
(C) the mass of the object, relative to a stationary observer, approaches infinity[2].

I will now attempt to show, under the restrictions that my neighbor laid down, that at least (A) and (B) must be true. Then I will quote an argument, under the restrictions, for the truth of (C), as made by a physics graduate student .

1. Miller, Jr., Franklin, *College Physics*, 4th ed., Harcourt Brace Jovanovich, Inc., N.Y., 1977, p. 154.
2. "There are, of course, other effects besides A, B, C: Einstein's original paper, for instance, was titled "On the Electrodynamics of Moving Bodies", and focused on how electromagnetic fields transform under change of reference frame." — E. B.

Plausibility Argument that Time Slows Down...[1]

Suppose we have a rocket ship that is 186,000 miles long. Aboard this ship are two mirrored plates that are parallel to the direction of movement when the ship is moving. A pulse of light travels from some fixed point on one plate, *perpendicularly* to the mirrored surface on the other plate, from which it is reflected back to the first plate. Each moment the pulse returns to the first plate, we consider to be a “tick” of the clock.

Assume now that our ship is at a steady speed that is near the speed of light. Since by our assumption the speed of light is constant throughout the universe, it will clearly take the light pulse longer to travel to the mirrored surface and back, because the pulse will travel a diagonal path due to the forward movement of the ship, instead of the simple vertical path it would travel if the ship were stationary. So the clock aboard the ship will run slower relative to an observer who is stationary relative to the motion of the ship. A shipboard observer will detect no change in the clock’s speed.

Can we ignore the mirrored-surfaces model, and say that the reason that time slows down is simply that, if a speed s of something, e.g., light, is d/t, where d is distance and t is time, and we decrease d by some amount, that is, multiply it by a real number r, where $0 < r < 1$, then to keep s the same, we must also multiply t by r, i.e., we must slow down time?

Plausibility Argument that Object Lengths Shrink With Speed

Imagine that there is a light source at the rear end of the rocket ship. Assume the ship is parked in front of the on-the-ground observer’s lab so that the midpoint of the ship is directly in front of the observer. Assume that when the light pulse is emitted down the length of the ship, another light pulse is emitted from the rear of the ship toward the on-the-ground observer. Then, when the pulse reaches the other end of the ship, a light pulse is simultaneously emitted from the front of the ship toward the on-the-ground observer.

When the ship and the on-the-ground observer are both stationary relative to each other, the observer must observe at least a 1-second interval between the sending of the pulse from the rear of the ship and the arrival of the pulse at the front.

Now assume that the ship goes past the on-the-ground observer at close to the speed of light, and that when the ship is exactly at the position it was in its previous stationary position, a light pulse is sent from the rear of the ship toward the front, and a second pulse is sent toward the on-the-ground observer as before. And similarly when the pulse reaches the front of the ship.

But the on-the-ground observer now reasons as follows: “It will take the pulse much longer to reach the front of the ship because of the speed of the ship. (If the ship were going at the speed of light, the pulse would never get to the front.) Therefore I would measure the speed of light as slower than before. But this contradicts the fact that the speed of light is always measured to be the same, namely, 186,000 miles per second. So the only possibility is that the length of the ship decreases.

A Seeming Contradiction

Now suppose that the above rocket ship is traveling at close to the speed of light. Suppose that no light pulse from the rear of the ship is fired, but that a pulse *from the front* is fired toward the

1. This argument is derived from a display at the Albert Einstein exhibit at the New York Museum of Natural History, Feb., 2003, but appears in many popular treatments of relativity.

rear. This pulse travels much *faster*, relative to the ship, than it would have if the ship were parked in front of, say, a laboratory.

But for an observer aboard the ship, the speed of light is constant. So it must be that the length of the ship *increases* and the on-board clock *speeds up*, so that, for the on-board observer, the speed of light remains constant. He observes no change in the length of the ship and the speed of his clock. An external observer, however, would observe the *increase* in the length of the ship.

We seem to have a contradiction between this case and the case that a light pulse is fired from the rear of the ship. How can it be explained?

A physicist strongly disagreed that we seem to have a contradiction, but I found his argument incomprehensible. Like all physicists I have communicated with, he seemed not to understand the concept of closing speed. Let me explain it briefly.

The Concept of Closing Speed

Suppose car A is traveling at 100 mph down a straight level road.

Suppose car B is heading directly toward car A on the same road at 100 mph.

Then their closing speed is 200 mph. Note that the closing speed is different from the speed of either car.

If car B is stopped some distance directly in front of car A, then their closing speed is 100 mph.

Suppose car B is following directly *behind* car A. Suppose car A is traveling at 100 mph. Then:

If car B is traveling at 200 mph, their closing speed is 100 mph.

If car B is traveling at 100 mph, their closing speed is 0 mph.

If car B is traveling at less than 100 mph, their closing speed is negative, meaning that car B will never catch up to car A.

Another explanation is the following:

Suppose there is a train, call it train A, on a straight, level track. The train is 100 feet long. Suppose there is a duplicate train, call it train B, on an adjacent straight, level parallel track. Suppose the front of train A is adjacent to the front of train B. Now suppose that train A instantaneously starts moving forward at 100 feet per minute. Train B remains motionless. Then in one minute, the front of train A will be adjacent to the rear of train B.

Now suppose instead that the moment that train A starts moving forward at 100 feet per minute, the front of the train B starts moving (in the opposite direction from the movement of train A) at 100 feet per minute. Then in *half a minute* the front of train A will be adjacent to the rear of train B. The reason is that in half a minute, the movement of train B reduced to 50 feet the distance that the front of train A had to travel in order to be adjacent to the rear of train B, and train A travels 50 feet in half a minute.

Observe that the speed of train A *was the same in both cases*. The *closing speed* of the two trains, however, increased from 100 feet per minute to 200 feet per minute.

The physicists I have communicated with believed that the speed of train A had to increase to 200 feet per minute, which is clearly nonsense.

The physicist to whom I sent the above scenario also seemed to believe that, even though it is meaningful to say that the length of the ship shrinks (as measured by an external observer) and the on-board clock slows down when the light pulse is fired from the rear of the ship to the front, the length of the ship does not increase (as measured by an external observer) when the light pulse is fired from the front of the ship to the rear.

I will only read comments on this that explain, step by step, what the author of the comments believes an external observer sees and does in each case. We know that an internal observer will see no change in ship length or clock speed in either case.

Plausibility Argument that Mass Increases With Speed

"I don't know of a similarly intuitive derivation of how energy/mass depends on velocity, but at least the following makes sense: a massive particle cannot go as fast as the speed of light — it would take an infinte force to accelerate a massive particle to the speed of light — so a particle's mass must approach infinity as it accelerates to the speed of light." — E. B.

Some Things That Bother Me About Events Leading to Special Relativity
The Assumed Need for an Ether in the First Place

In the 19th century, light was believed to share properties of both waves and particles. Physicists found this troubling. They felt that a medium, called the "ether", must exist to "support" light waves, just as a medium (namely, molecules) exists to support waves in air and water.

Why didn't physicists just assume that particles of light were sufficient to support light waves, just as particles (molecules) of air and water are sufficient to support waves in air and water?

Suppose Michelson Had Simply Declared His Experiment Showed There Is No Ether?

If I had been Michelson, I would have concluded from the surprising results of my and Morley's experiment — namely, the results that showed that motion through the supposedly-existing ether has no effect on light — that the ether doesn't exist! Suppose that Lorentz and FitzGerald and all other leading physicists had agreed with me. Then there would have been no need for Lorentz to come up with the Transformation he did to show that objects shrink with speed.

Then that would have left Einstein with the fact (established by the Maxwell equations) that the speed of light is constant. From there, he would have been able to deduce the relativity of simultaneity via his famed railway train example .

Now it is inevitable that physicists would have wondered what the relationship is between the values of x, y, z, t in different inertial frames, e.g., one relatively stationary frame, and another relatively moving frame. Perhaps then physicists would have derived the transformation equations I have derived in "Appendix C — Is There an Alternative to the Lorentz Transformation?" on page 67

These equations, like the Lorentz Transformation, show that length (in the direction of travel) shrinks with speed, and time slows down with speed. So these remarkable facts about Special Relativity would not have been lost due to the alternative transformaion.

Some Things That Bother Me About Special Relativity
Measuring the Speed of Light

If we ask how exactly we would go about measuring the speed of the light pulse as the rocket ship in “Plausibility Argument that Object Lengths Shrink With Speed” on page 6 goes past at close to the speed of light, we find that the answer is not so simple. The reason is that we must take into account the time it takes for the pulses that are sent to the observer, to reach the observer. We must also have a way of knowing exactly how far the center of the ship was from the observer when at least one of these pulses was sent.

Therefore I ask physicists to describe exactly how the speed of light in the rocket ship (a relatively moving frame) would be measured from a stationary point (a relatively non-moving frame).

(A physicist has written to me stating the following: “An observer never makes measurements in someone else’s frame.”)

The Constancy of the Speed of Light

Thinking about these matters, it seems unavoidable to ask: *what exactly does it mean to say that the speed of light is constant in a universe in which rulers shrink and clocks slow down in an object whose increasing speed approaches that of light*?

In trying to answer this question, it might occur to us to ask what exactly we mean when we say that the length of objects shrinks as objects approach the speed of light. In “Plausibility Argument that Object Lengths Shrink With Speed” on page 6 we used a deductive argument and concluded that the length of the ship must shrink. But suppose, in accordance with the basic assumption of Special Relativity, namely, that the speed of light is constant throughout the universe, we were to ask, Could we *measure* the shrinkage of the ship (as opposed to simply deducing that the shrinkage must occur)?

When the ship is traveling at near the speed of light, it takes longer than a second for a photon to travel the length of the ship, just as, if there is a train moving in a straight line at a constant speed v over flat ground, and we roll a ball at a speed $v + 5$ (relative to the ground) from the back of the train to the front, the ball is only traveling at 5 mph relative to the floor of the train. The smaller the difference between the train’s speed relative to the ground and the ball’s, the slower the ball travels relative to the floor of the train. (Here, the train is analogous to the rocket ship and the ball is analogous to a photon.)

But then it takes much more than one second for the ball to travel the length of the train.

Yet the speed of light must be constant, no matter under what circumstances it is measured. One way to assure this is if the length of the rocket ship shrinks so that a photon travels from the back of the ship to the front (a much shorter distance, because the ship has shrunk) in one second as measured by the observer aboard the ship.

But what about the external observer? He sees the photon having traveled only this short distance from the back of the ship to the front. Since the speed of light is constant, he computes that the time to travel that short distance is only a fraction of a second. But he sees that the observer’s clock aboard the ship, reads one second for the proton to travel that same distance. Does he say that the observer’s clock has “slowed down”? Is he correct in reasoning that if the observer’s clock takes one second to cover the same amount of time that his clock covers in a fraction of a second, then the observer’s clock has slowed down?

On the other hand, if *instead of* the ship shrinking, suppose that the observer’s clock slows down sufficiently so that it measures one second from the time that the photon leaves the back of the ship, to the time it reaches the front (because the photon is moving at a snail’s pace).

So it appears that the shrinking ship *or* the slowing of the observer's clock (one or the other, but not both), are two ways to enable the observer to measure the correct speed of light even though the speed of the ship increases.

We ask again: what exactly does it mean to speak of a fixed speed (namely, that of light) in a world (universe) in which length decreases and time slows down in any object as it approaches the speed of light? (For example, in an object whose length is being used to measure a fixed speed?) Contrast this with the following: we have a pistol that fires bullets whose speed is always the same. So no matter where we go on Earth, and no matter how we measure the speed, we always come up with the same result — a result that is given by the formula: speed of bullet = (distance that bullet travels)/(time it takes to travel that distance), because no matter where we go on Earth, a centimeter, an inch, a foot, a meter, etc. always has a constant length. This seems perfectly straightforward. The speed-of-light case does not seem to be.

Why the Speed of Light is Constant

For some reason, physicists — and even Einstein, it seems — do not explain the obvious reason why the speed of light is a constant. It is because of a simple definition:

(the wave velocity of light) = (the speed c of light)= $(f)(\lambda)$, where f is the frequency of a given light ray, and λ is the wavelength. See, e.g., Giancoli's *Physics*[1], pp. 288, 529.

The speed of light is approximately 186,000 miles/sec = approximately 300,000 km/sec.

On the Shrinkage of Objects With Speed

Suppose there is a space ship. I know its length when it is parked in front of my laboratory. I now arrange to have the ship pass from East to West at a known distance from my laboratory, while I am facing North with a camera. The ship's speed is close to the speed of light, but I have programmed my camera to take a picture of it at the moment it is directly in front of me.

Since I know the distance to the ship when the photo was taken, I can do a little trigonometry on the photo, and determine the length of the ship when it passed.

Would I compute the length to be less than the length when it was parked in front of my laboratory? A physicist has written me that the answer is Yes. However, the following is an editor's footnote on p. 13 of Lillian Lieber's s *The Einstein Theory of Relativity*[2]:

> Until 1959 the Lorentz contraction of a sphere into an ellipsoid was believed to appear just as Prof. Lieber describes. That year, R. Penrose and J. Terrell independently showed that a photograph taken of a very rapidly moving sphere would show, surprisingly, a sphere. R. Penrose, "The Apparent Shape of a Relativistically Moving Sphere", Proc. Camb. Philos. Soc. 55 (1959); J. Terrell, "Invisibility of the Lorentz Transformation," Phys. Rev. 116 (1959): 1041-1045.

---

1. Giancoli, Douglas, *Physics*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980.
2. Paul Dry Books, Philadelphia,, PA, 2008.

Robert J. Buenker in his *Relativity Contradictions Unveiled*[1], states: “In order to satisfy the light-speed constancy postulate, it is essential that lengths expand when clocks slow down, and in exactly the same proportion.”[2]

That is, since distance = rate × time, if we let rate = *c*, the speed of light, we have *c* = distance/time. Now *c* is constant, so if time slows down (i.e., increases in value, so that, e.g., 1 second now becomes 10 seconds), then the only way for *c* to remain constant is for distance to *increase*. But this is in direct contradiction to the basic precept of Special Relativity that lengths *decrease* as an object’s speed approaches the speed of light.

Permanence of Slowing of Clocks, Non-Permanence of Shrinking of Objects

We are familiar, from science fiction literature and films, with the phenomenon of a person who has been on a space trip at speeds close to the speed of light, having aged much less, when he returns to earth, than those he left behind. But why is it that the person, and objects he had with him, do not also show the shrinkage that took place when they were near the speed of light? Why does this shrinkage “go away”, whereas the effect of the slowing of clocks, remains?

Why is it that the masses of the person and the objects are not considerably greater on return, than they were when they left?

The Relativity of Simultaneity

Early in 2007, I wrote the following email to the physicist E.B.:

> The following occurred to me re the standard argument against simultaneity being absolute in the universe. I am referring here to the argument about lightning bolts striking in front of and in back of a train, and there being an observer on a hill equidistant from the points at which the lightning bolts strike, another observer at the front of the train, another at the back.
>
> I don’t dispute that the bolts will appear to strike simultaneously for the observer on the hill, or that the bolt in front of the train will appear to strike before the one at the back for the observer in the front of the train, etc. But I dispute that the observers must necessarily be so naive as to go solely by what they see!
>
> Suppose I’m an astronomer. One night, looking through my telescope, I see two supernovas that appear to occur simultaneously. My reaction would be to say to myself: “How interesting! I have just become aware of an infinity of possibilities regarding two supernovas. The only fact I can assert is that the light from two supernovas reached me at the same time. The possibilities are all the ways this could happen: one supernova could be close, so that the actual explosion took place only a few years ago, while the other supernova could be far, so that its actual explosion took place many years ago, but the light only reached me now. A table, or a graph, representing all these possibilities, is all I can hope to come up with, barring further information.”
>
> Similarly, in Special Relativity, every observation should result in a *table* that represents all the possible ways the observed event could have occurred. So, for example, the guy in the

1. Apeiron, Montreal, 2014
2. ibid., back cover summary of p. 34

front of the train should say to himself, “Hmmm. The lightning bolt that struck in front of the train appears to have struck before the bolt that struck in back of the train. Now let me try to represent all the ways this could have happened...” and then, since he knows the speed of the train and the length of the train and the speed of light and the observed time lapse between the two bolts, one of the possibilities in his table will be that in fact the bolts struck simultaneously, as perceived by the guy on the hill.

E.B. replied as follows:

Your comment is well within the spirit of relativity — an observer realizes that what he sees “naively” is only one of an infinity of possible descriptions of the event, and he knows how to adjust his “naive” description so to be consistent with any of the others. the only problem with what you say is in implying that one description is more “naive” than another — i.e. that there is one “correct” description — in fact, all descriptions are equally valid and equivalent.

But I have trouble accepting E. B.’s statement. At least part of his reply is based on the basic assumption of Special Relativity that it is not possible for an occupant of an inertial frame to know if the frame is moving (at a steady rate) of if it is motionless.

Precept Regarding the Determination of Frame Movement From Within a Frame

The Precept says, in full, that it is impossible to determine, by a mechanical experiment within an inertial frame, whether or not the frame is moving.

I interpret “mechanical experiment” to mean “experiment involving objects having mass.” I do not dispute this. However, I believe it is possible to determine, by an experiment *involving only photons*, which have no mass, if an inertial frame is moving.

I must make a special request to physics graduate students and physicists to read and think about what I have actually written, as opposed to what, at first glance, they think I have written. I also must make clear that I have no interest in proving that Einstein was wrong. My only interest is in finding out if the argument presented in this section is valid, and if it is corroborated by the Test described at the end of the section.

Conjecture on Detecting Whether or Not an Inertial Frame is Moving.

Preliminaries

I will assume that the reader is familiar with Einstein’s moving-train model to show the relativity of simultaneity, that is, the fact that two events may appear to occur simultaneously to one observer, but not to another. This model is described in virtually all popularizations and also in more in-depth treatises of Special Relativity. A translation of Einstein’s description of his model can be found on pp. 25-26 of his book, *Relativity*[1].

To quickly review: in the model, a lightning bolt strikes at a distance d in front of a train moving in a straight line on a level track at constant speed v, at the same moment as another lightning bolt at a distance d behind the train also strikes. “At the same moment” means relative to an

1. translation by Robert W. Lawson, Prometheus Books, N.Y., 1995.

external observer on a nearby embankment who is sitting on a line perpendicular to the center of the train at the moment the bolts strike. This observer sees the bolts as occurring simultaneously because the speed of light is constant, and the distance from each bolt to the observer is the same.

However, an observer sitting in the middle of the train sees the bolt in front as occurring before the bolt behind. The reason for this difference is that because the photons from the front of the train are moving toward the middle of the train, and the middle of the train is moving toward the photons from the front, the time for the photons to reach the middle of the train is less than the time for the photons from behind the train to reach the middle.

Some readers have trouble understanding this difference in arrival times at the middle of the train, so let me express it in terms of closing speeds. In the case of Einstein's train model, the closing speed of the photons from the front of the train, and the middle of the train, is $c + v$, where c is the speed of light. But the closing speed of the photons from behind the train, and the middle of the train, is $c - v$. Since the distance to be traveled by the photons in each case, is the same, namely, $d + (½)r$, where r is the length of the train, the photons from the front of the train will reach the middle, hence reach the observer, before the photons from behind the train.

Around 2015 I began asking myself, Suppose that instead of the two lightning flashes, there had been simultaneous flashes of light from lightsources at the end of each of two metal rods projecting equal distances d from the front and rear of the train. (It is a well-known fact that clocks (timers) can be synchronized within an inertial frame — in this case, to set off the flashes simultaneously.[1]) What would the observer in the middle of the train have seen?

It seemed to me clear that the observer would have seen exactly what he saw in Einstein's original model.

But I realized that *that implied that it is possible, from within an inertial frame (the train), to determine if the frame is moving, which is an exception to one of the basic precepts of Special Relativity!*

I devised a possible proof of my Conjecture. It follows. (Note: A much shorter, and, I believe, clearer, version of the First Possible Proof is given under "Second Possible Proof of Conjecture" on page 14.)

First Possible Proof of Assertion

The possible proof is presented in an informal style, though I believe it is logically valid.

1. The only difference between my train model, and Einstein's original one, is that the flashes in mine come from two metal rods projecting from the front and rear of the train, whereas in Einstein's they come from two lightning bolts striking the ground. Otherwise everything is the same.

---

1. *Proof*:
Assume there are two identical timing devices, A and B — identical in parts, assembled in one and the same shop.
Assume that in that same shop, the starting time in each device is set at 0 min..
Assume that the Start button on each device is simultaneously pressed. The devices are observed ticking at the same rate, and are observed to reach the successive time figures (e.g., 1 min., 2 min., 3 min., …) at the same time.
One of those running timing devices is then placed next to and connected to the front lightsource, and the other running timing device is placed next to and connected to the rear lightsource.
The same time figure on each device can then be set to cause the lightsources to flash simultaneously at that time.

The reader may find it helpful to imagine the two experiments being conducted in parallel, meaning, in particular, that the photons are released by the lightning bolts at the same moment that they are released from the lightsources on the metal rods.

2. The only way that the observer at the middle of the train could see something different in my model, would be if the speed of the photons as a result of their having been emitted from a moving source, was different from the speed of the photons in Einstein's model, which are emitted from fixed sources.

3. But the photons have one and only one speed. In fact, *there is no difference in the behavior of the two models from the moment after the photons are released*. The constancy of the behavior of photons is another Basic Precept of Special Relativity. Therefore, to argue that it is different in the case of the flashes from the two rods, is to argue against that Basic Precept of Special Relativity.

So the observer sees what he saw in Einstein's model. Therefore it is possible to tell, from within a moving inertial frame, that the frame is moving. □

Note: If, instead of photons, we had used billiard balls rolling along a track inside the train and running the length of the train, it would have been possible to get the balls to arrive at the observer at the same time, All we would have had to do was roll the balls at the same speed relative to the floor of the train.

If I am right about the passenger in the modified model seeing exactly the same thing he saw in the unmodified model, then all I have really shown is that:

- *No tests performed solely with objects that have mass, can reveal, to persons in an inertial frame, if the frame is moving or not relative to an external observer.*

Second Possible Proof of Conjecture

1. Assume an inertial frame has a lightsource in front of it and a lightsource in the rear.

2. Assume there are light detectors in the middle of the frame, one to detect light from the front lightsource, one to detect light from the rear one.

3. Assume the lightsources are programmed to emit light at the same time, and that timers in the light detectors are programmed to record when the light from each source reaches the detector.

4. If the detector for the front lightsource receives light before the detector for the rear lightsource, then the frame is moving (in a line in the forward direction). If not, the frame is not moving.

Responses of Physicists Who Were Told About the Possible Exception to the Precept

I sent emails to at least seven physicists at various universities. In these emails, I described my modification of Einstein's model, and stated that if the observer in the middle of the train saw what he saw in Einstein's model, namely, the flash from the front before the flash from the rear, then that would imply an exception to a basic precept of Special Relativity (see above). I had no idea the reaction would be what it was.

Several physicists said in no uncertain terms that any suggestion that there might be an exception to a basic precept of Special Relativity could only come from a crackpot.

Two said that the error in my thinking can only become clear via higher mathematics. The error could not be explained at the photon level of Einstein's train model.

But several other physicists said that higher math is not required. I am simply confusing the various inertial frames involved. (But I am using only one inertial frame, namely, that of the train.)

Others said that the error in my thinking is well-known, and is explained in most textbooks. (But I have looked at six textbooks, plus a variety of Google articles, and found nothing about my version of Einstein's train model.)

One physicist said he performs "frequently" in the lab in his apartment, an equivalent experiment and always finds that the photons from the front and the back arrive at the observer at the same time. (But his apartment is not moving at a constant speed *v*. When I questioned his experiments, he said I was never, ever to write him another email.)

Another physicist said that my error is my claim that light has two speeds, $c + v$ and $c - v$, whereas it is known that light has only one speed. But I make no such claim: the two speeds are *closing speeds* of photons and the observer, as is clearly explained. The speed of light, *c*, is the same in both. I pointed this out to the physicist. He ignored my argument. He then said that I was merely re-inventing "Galilean invariance". This is the principle, set forth by Galileo in the early 1600s, that it is not possible to tell, from within a closed space moving at a constant speed, if the space is moving or not. Galileo's closed space was the interior of a ship moving at a constant speed within completely calm waters.

When I asked him to explain how I was doing that, he said that my refusal to accept his criticism as correct, whether or not I understood it, was "breathtakingly arrogant" and that I was never to write him again.

Possible Explanation for Physicists' Criticisms Having Little To Do With What I Actually Wrote

No physicist ever bothered to explain where the error in my assertion lay. It seemed clear that once a physicist knew that I was setting forth a possible exception to a basic precept of Special Relativity, he felt no need to actually read and think about the modified version of Einstein's model that I presented. Since there could be no exceptions to the basic precepts of Special Relativity, virtually anything could serve as a valid criticism. For example, I was amazed to learn that, after some six months of communicating with me about my assertion, and presumably thinking about it, one physicist revealed that he wasn't sure if the train in my version of the model, was moving!

How I Got the Physicists To Agree I Was Right Without Admitting It

In June, 2017, I sent the following email to 12 physicists at several universities:

Prof. ...:

You are no doubt familiar with the train model that Einstein used to establish the relativity of simultaneity.

Suppose that instead of the two lightning flashes, there had been simultaneous flashes of light from lightsources at the end of each of two metal rods projecting equal distances from the front and rear of the train. (It is a well-known fact that clocks (timers) can be synchronized within an inertial frame -- in this case, to set off the flashes simultaneously.) What would the observer in the middle of the train have seen?

If I do not hear from you, I will assume that you believe the observer would have seen what he saw in Einstein's original model, namely, the flash from the front of the train before the flash from the rear.

If I do hear from you, please be assured that I guarantee complete confidentiality in all communications, and that I will not argue with anything you say.

Regards,

-- Peter Schorer

Note that I did not mention the possible exception to the basic precept of Special Relativity.

Ten out of 12 of the physicists I wrote to did not reply, so I assume they agreed I was right. Two did reply. One of these replies I didn't understand, in the other one the physicist said that the observer would see both flashes at the same time. When I asked him how he arrived at that conclusion, he said an explanation wasn't possible via email(!).

I then wrote to two of the physicists who had not replied, informing them of the implication, and asking for their thoughts. I told them I guaranteed complete confidentiality. Neither replied.

One morning at Peet's Coffee and Tea store near Rick & Ann's restaurant, in Berkeley, CA, where I often went for breakfast, I was sitting at the high, narrow table facing the window. A man came up, asked me what I was studying. I said mathematics. I asked him if he was a mathematician. He said no, a physicist. He worked at a major lab in the area. I immediately told him I would welcome his answer to a physics question I had. I described my modification of Einstein's model, then asked him what the observer would see, and without a moment's hesitation he said he would see the same thing as before. I then told him what this meant for a basic precept of Spacial Relativity. He smiled, and nodded thoughtfully. We shook hands, I told him my name, he told me his, which I immediately forgot, and he left. I forgot to ask him for an email address, so I have been unable to contact him again.

But in any case, I now believe I may have discovered something important.

A Test of the Validity of My Assertion

Given the accuracy of today's atomic clocks, it seems that a test of my claim could be performed aboard a long jetliner (or a long train!) with lights flashed simultaneously from the front to the center and from the rear to the center. A forward-facing detector at the center would record the elapsed time since the front light was flashed, and a backward-facing detector would record the elapsed time since the rear light was flashed. If the time for the forward-facing detector were less than the time for the backward-facing detector, then we would have reason to believe that my assertion might be valid.

Light As a Measuring Tool in Inertial Frames

Writers on Special Relativity spend considerable time on the meaning of measurement in an inertial frame. Lengths are stated as being measured by end-to-end placement of a ruler, clocks are stated as having hands, and the problem of synchronization of the clocks is then dealt with at length. Why cannot a finite set of clocks be synchronized by (1) the clocks all being identical mechanically, and (2) the clocks all being set at the same time when the clocks are in the same place, and are then moved to where they are to register the time?

But why is it necessary to employ such crude means, when we have a tool that can be regarded as the best possible, namely, *light*. The speed of light is a constant, regardless of the movement of the frame in which it is emitted. It would seem that, from the basic physical fact that distance (x) = speed (v) times time (t), we can let the speed be that (c) of light. Then to find the distance between two points in an inertial frame, we need only determine how long it takes a light pulse to travel from one to the other. To determine the speed of an object in a frame, we simply divide the distance it travels between two points by the time it takes light to traverse the distance.

There is, of course, something circular about this. To measure a distance, we need to know a time. To measure a time, we need to know a distance, etc. The question is, can this apparent circularity be overcome?

On the other hand, in the traditional treatments involving rulers and clocks, we must ask: why do we believe that the ruler lengths are what we believe them to be? Why do we believe that the clock ticks are what we believe them to be? In other words, why is there no circularity in the traditional treatments?

Now that we are in a time when atomic clocks and radio clocks exist, it seems that measurements in, and between, inertial frames would far better be expressed in terms of these devices.

Presentations of Special- and General-Relativity

Several Popularizations

As any reader knows who looks at the physics section of a college bookstore, or of any good used-book store, or who probes the Internet, there have been very many attempts to present Special- and/or General-Relativity at various levels of detail.

As far as popularizations are concerned, among the best I have come across is Martin Gardner's *The Relativity Explosion*[1]. The main reason for the book's quality is, I'm sure, Gardner's long experience in writing for a wide, technically-literate audience, which is no doubt also why he chose to include so many illustrations. Bertrand Russell's celebrated *The ABC of Relativity*, is, in my opinion, not as good as Gardner's work. Although Russell was one of the best expositors of technical concepts in the 20th century, his geometric explanations of the space-time interval, and of the derivation of the Lorentz Transformation, are an embarrassment (for him). I challenge any reader to write down a clear explanation of the latter explanation. These explanations make one think that Russell knew far less about the formal presentation of mathematical proofs than we assumed.

1. Vintage Books, N.Y., 1976.

Lillian Lieber wrote an excellent text, *The Einstein Theory of Relativity*[1]. (Unlike the overwhelming majority of physics and mathematics textbooks, this one gives the justification for *each* statement.) However, the first part of her derivation of the Lorentz Transformation is, at least for me, baffling.

My attempt to derive the Lorentz Transformation using the simplest means possible is given in "Appendix C — Is There an Alternative to the Lorentz Transformation?" on page 67.

Towards a Mathematically-Rigorous Presentation of Special Relativity

A major problem with popularizations, including those that use some mathematics, is that the authors believe that the subject can be made less intimidating by explaining as much as possible in prose. But this is simply not true — we can call it the Popularizer's Fallacy — once the discussion gets down to basic operations such as the synchronizing of clocks, the measurement of lengths and times in inertial frames, etc. The reader feels, on the one hand, that he certainly should be able to understand matters as simple as clocks and the use of unit-length measuring rods, but on the other hand, if he allows questions about what he has read to emerge in his mind, he finds that he is not at all sure how to answer them.

After a good deal of struggle trying to understand Special Relativity, I have come to believe that the only remedy is a mathematically-rigorous presentation of the subject *using only the level of mathematics that is necessary to understand, and prove, the Lorentz Transformation*. I have never come across one. There are, of course, sophisticated presentations of Special Relativity that use mathematics considerably more advanced than that required for the Transformation. I am not interested in these as long as popularizations continue to imply that Special Relativity can be understood at the Transformation level.

The kind of presentation of the Theory of Special Relativity that I have in mind would have the following characteristics:

Characteristics of the Presentation

• clear statement by the author, at the beginning, of the minimum knowledge he is assuming among readers. The presentation must then provide proofs for all facts not part of this minimum knowledge.

• structured proofs (analogous to structured programs in computer science) wherever appropriate.

• justification (either explicit or by reference) for each and every statement, whether the statement is in a proof or not. (See the chapter "Proofs" in William Curtis's *How to Improve Your Math Grades* on occampress.com.)

• complete, thoroughly cross-referenced index, including an index of symbols and of frequently-occurring expressions.

---

1. Paul Dry Books, Philadelphia,, PA, 2008.

• testing of the finished presentation on randomly-selected readers having the minimum knowledge the author specified.

A rigorous presentation would all but eliminate the need for the endless *explanations* and *plausibility arguments* and *examples-designed-to-convince-the-reader* and *long-winded prose* that characterize so much of the commonly available literature on Special Relativity[1]. The answers to most questions about the subject would be simply a reference to one or more definitions and/or proofs in the presentation.

Questions that Definitions in the Presentation Must Answer

I feel that the all the following questions must be answered in the presentation. If any of the questions are irrelevant, then the reason why must be stated.

• *Speed of Light* — what exactly does it mean to say that the speed of light is constant in a universe in which rulers shrink and clocks slow down in an object whose increasing speed approaches that of light? (See "Why the Speed of Light is Constant" on page 10.)

• *Inertial Frame* —

What is the formal definition of an inertial frame?

Does a frame have boundaries, and if so what are they? If not, how do we know where one frame begins and the other ends?

Can a frame occupy an infinite space? If not, then why are the two three-dimensional sets of Cartesian coordinates that are routinely diagrammed in derivations of the Lorentz Transformation, called "frames"?

Can one frame be inside another frame?

Can an observer in one frame observe events in another frame? (A physicist has told me that the answer is No. But in Einstein's *Relativity*[2], p. 117, we read "In order to see how the points of the x'-axis appear as viewed from K, we only require to take a 'snapshot' of K' from K." (K and K' are separate inertial frames.))

Is a frame a single, continuous, connected entity?

---

1. Robert J. Buenker's book, *Relativity Contradictions Unveiled,* which was mentioned in the section, "The Constancy of the Speed of Light" on page 9, is an example of a controversial work that suffers greatly from the fact that its claims are not presented in the context of a mathematically-rigorous presentation of Special Relativity. As a result, the few physicists who deign to look into the book (anyone who dares to question Special Relativity is regarded by physicists as a crackpot) will be able to come up with counterarguments, counterexplanations, counterexamples, to which, I am sure, Buenker will reply with more of the same. Each side will become convinced that it is right and the other is wrong, the whole exchange being, in my opinion, a waste of time that could largely have been avoided by a mathematically-rigorous presentation.
2. Einstein, Albert, *Relativity*, Prometheus Books, Amherst, N.Y., 1995.

Does a frame need to have an observer inside it?
If so, then where is the observer in a frame located?

How does an observer recognize an event in his own frame, and in another frame? Are there speed-of-light considerations that must be taken into account if distances between observer and event are sufficiently large, even in one frame?

What are all the tasks that can be performed by an observer in his frame, and in another frame? Is it even possible for an observer in frame F_1 to perform tasks in another frame F_2? (See "Appendix B — On Basic Tasks in Special and General Relativity" on page 63.)

• *Coordinate Systems* —
Is the coordinate system in each frame always the same? If so, is it always Cartesian coordinates?

Where is the origin of the coordinate system located in each frame? Near the "front" of the frame? Anywhere at all?

Is a frame more than just a set of coordinates, and, if it is, what else is contained in a frame?

In the Lorentz Transformation, what exactly do the terms, x, y, z, t, and x', y', z', t', represent?

What determines $t = 0$, $t' = 0$? Do the terms always pertain to a specific event, e.g., a lightning flash, or do they pertain to a continuous sequence of events, e.g., the outward movement of the spherical wavefront from a flash of light?

To put the questions in another way: if x, y, z, t, are the coordinates in the relatively stationary frame F_1, and x', y', z', t' are coordinates in a frame F_2 moving at a speed v relative to F_1, what exactly does the first equation of the Lorentz Transformation say *to the observer in* F_2? Does it say, "If an event in *your* frame (F_2) occurs that has the coordinates x, y, z, t in the *other* frame (F_1), then the coordinate x' in your frame, of the event, will be given by

$$x' = \beta(x - vt)$$

where

$$\beta = c/(\sqrt{c^2 - v^2})$$

And does the last equation of the Lorentz Transformation say *to the observer in* F_2? "If an event in *your* frame (F_2) occurs that has the coordinates x, y, z, t in the *other* frame (F_1), then the coordinate t' in your frame, of the event, will be given by

$$t' = \beta\left(t - \frac{vx}{c^2}\right)$$

This would seem to imply that the observer in F_2 can see into F_1, and that, in particular, he can somehow know not only what the event is, but also what the values x, y, z, t are. Does the speed of light become relevant here? Suppose the frames are several light years apart?

• *Speeds* — are the following statements correct?

(1) If an object u in a frame is moving at a velocity v relative to that frame, and the frame is moving at a velocity w relative to some other, relatively stationary frame, then the total velocity of r relative to the relatively stationary frame, is $v + w$ — as long as the object u is not a photon. For, the speed of the photon is c, the speed of light, which is a constant. But nothing can travel faster than c, and so it would appear that w must be 0, which it is not. A resolution of this seeming contradiction is given under "Another Possible New Precept for Special Relativity" on page 22.

(2) If two objects in a frame are moving toward each other, one at a speed v, the other at a speed w, then the closing speed of the two objects, as observed from another frame, is $v + w$. If v and w are both greater than 0, then the closing speed is greater than v and greater than w. However the speeds v, w are unchanged. Thus, if $v = c$, the speed of light, which is a constant, we do *not* have a contradiction here, since closing speed is not the speed of an object.

• *Shrinkage of Lengths in Direction of Speed of Moving Object* — do the lengths "really" shrink, or is the shrinkage just an optical illusion? If they really shrink, how is this possible given the molecular and atomic forces in the molecules and atoms composing the object?

• *Slowing Down of Time in a Moving Object* — does the effect of time's slowing down (time dilation) with an object's high speed, remain after the speed has slowed down? Apparently the answer is Yes (see, e.g.,the film *Close Encounters of the Third Kind*, or The Twins Paradox in special relativity treatises.)

Assume a spaceship parked before us and pointing in a direction perpendicular to a line extending in front of us.

Suppose it had a window in its side through which we could see a light that flashed, say, each second (the flashes constituted the ticks of an onboard clock).

Now assume that the ship went out into space and gained speed to, say, half the speed of light, and then flew past us, at one point being at the same location as it was when we observed the clock ticks when it was parked before us.

Suppose we made a film or video of its passing by. Would the flashes now appear to be occurring much slower, i.e., would they occur, say, every 10 of our seconds?

Another Possible New Precept for Special Relativity

Assume a train is moving in a straight line at a speed v. Assume a ball is rolled forward, inside the train, at a speed w, where $w < v$. Then the speed of the ball relative to an observer in a frame that is stationary relative to the train, is $w + v$.

Now assume a spaceship is moving in a straight line at a speed v that is close to the speed of light. Assume a beam of light is flashed from inside the back of the ship, directly towards the front of the ship. The speed of a photon in the beam of light is c. However the speed of the photon relative to an observer in a frame that is stationary relative to the spaceship, cannot be $c + v$, because nothing can travel at a speed greater than that of light. So $c + v = c$, implying that the speed of the spaceship is zero, contradicting our initial statement as to its speed.

We are told, in Special Relativity, to ignore arguments like this, and simply to accept, as a basic precept, that nothing can travel faster than light. However, why shouldn't we instead say that $v = 0$ simply implies that *the speed of light can only be measured by an observer in the same frame as the beam of light in question*?

Thus all attempts thus far to measure the speed of light are perfectly legitimate, because the measurements have all taken place in the same frame in which the light or other electromagnetic wave occurs. Of course, the size of some of the frames is very large — millions of miles in the case of electromagnetic waves sent and received from earth to Mars rovers.

Clearly a rigorous presentation of Special Relativity must either accept the above implication as a new precept, or else must include a step-by-step description of how the speed of light, c, in a relatively moving frame, can be measured from a relatively stationary frame.

The Need for Annotated Editions of the Classic Early Papers on Relativity

The classic early papers are available in paperback, namely in Einstein, A., Lorentz, H. A., Minkowski, H., and Weyl, H., *The Principle of Relativity* (Dover Publications, Inc., Mineola, N.Y., 1952). The papers were written for the physicists of the time, i.e., of the early 20th century, and so the authors had a perfect right to make certain assumptions about what their audience could be assumed to know.

But for a person in the early 21st century who is neither a physics student nor a physicist, even though he may have had considerable mathematical training, there are many parts of these papers that are difficult, if not impossible, to understand. And yet, because of the importance of the papers, it would seem to be a worthwhile service to these readers (and possibly even to physics students and physicists!) to make available annotated editions of the papers. Part I of Einstein's first paper[1] on Special Relativity uses only high-school math (well, high-school math before the national decline in public schools). Nevertheless, even these parts are difficult. In fact, an advanced graduate student in physics was not able to penetrate the obscurities and omissions in this Part. Part II of the paper requires a knowledge of the elements of vector calculus.

"Appendix A — Obscurities and Omissions in Einstein's First Paper on Special Relativity" on page 59 contains a discussion of the sources of the difficulties in Parts I and II of Einstein's paper.

The kind of annotated edition I have in mind would be the equivalent of Martin Gardner's superb annotated edition of *Alice's Adventures in Wonderland*, namely, *The Annotated Alice* (New American Library, N.Y., 1960). Among other things, each edition would make clear what was already known to physicists of the time, what was controversial, and, most important, it would

1. "On the Electrodynamics of Moving Bodies", pp. 37-65, ibid.

explain the reasoning that is obscure — e.g., the material on the time function τ in section 3 of Einstein's first paper on Special Relativity (see "Appendix A — Obscurities and Omissions in Einstein's First Paper on Special Relativity" on page 59).

The goal of the annotated editions should be to decrease, as much as possible, the amount of time that a non-physics student or a non-physicist requires to understand the paper. I am confident that an annotated edition of Einstein's first paper on Special Relativity could reduce this time by a factor of *at least three*.

Needless to say, each edition should be tested on representative members of the intended audience before the edition is published. Unquestionably, the task-oriented approach described under "Environments Make Relativity Much Easier to Understand" on page 63, would help considerably in overcoming the obscurities that are inevitable (for non-physicists) with the discursive approach that is used in the papers.

Did Einstein Get the Idea of General Relativity From Bar Magnets?

It is well known that Einstein was fascinated by a magnetic compass that he received at age five from his father. But we may reasonably ask if he was not also later given, of if he did not also later acquire on his own, a bar magnet. If so, then he almost certainly learned to place the magnet on a piece of white paper, sprinkle iron filings around the magnet, and observe the curved patterns that the filings make as a result of the magnet's field. Perhaps, when he learned that the earth, and indeed all heavenly bodies, exerts a force (gravity) that attracts just as the magnet attracts, the idea might have occurred to him that gravity might curve space (actually, space-time) just as the magnet seemed to curve space and thus make the iron filings form their curved pattern.

Actually, the pattern of iron filings is not representative of the curvature of space-time around a massive object in space. Far more representative is the flow of water past a smooth, submerged rock in a stream.

Physics — Other

Entropy

Which is in a higher state of entropy: a stopped watch or a watch that is five minutes slow? Or are both in the same entropy state? Clearly, the behavior of both is equally "predictable".

When a working machine stops working, does its entropy increase?

"The 19th century classical notion of entropy as 'disorder' finally found microscopic basis through quantum mechanics in the early 20th century. Entropy is the logarithm of the number of available quantum mechanical states (at a fixed energy, number of particles, and volume, to keep things simple). When you're talking about, say, an ideal gas of particles, it's fairly easy to derive directly the 'disorder' notion of entropy starting from the microscopic definition. When you start talking about things like watches, the connection gets more tenuous due to computational complexity, and oftentimes one slips into the realm of metaphor. It sounds to me in your watch example that you are close to a metaphorical use of 'disorder'. The underlying quantum states don't care whether the watch is running slow or on time. That said, if there were a mechanical reason for a watch to start running slow, when it originally was running fast, there could be some entropic issues behind it. For instance, a battery that is using up its electrolytes is increasing its entropy all the time. Similar considerations apply for the working machine that stops: if it stops because some fuel source is used up, you're at a higher entropy state. If it stops because you flip a power switch and you can flip that switch back on, its entropy probably hasn't changed significantly." — E.B.

The Heisenberg Uncertainty Principle

The Heisenberg Uncertainty Principle says that you can't simultaneously measure both the position and velocity of an electron to any degree of accuracy you want. The more precise you want to measure the position, the less precisely you can measure the velocity. And vice versa.

Einstein was bothered greatly by this limitation in physics, and by the fact that probabilistic knowledge is the best one can hope for in certain areas of quantum mechanics. He apparently felt that probabilistic knowledge is incomplete knowledge, and reflects our inability to know how to penetrate to the core of the underlying reality.

But suppose we are limited to probabilistic knowledge when — to put it in computer terms — there simply aren't enough bits "to go around". In computer science, it is known that there are functions that can never be computed — not because we aren't smart enough to figure out how to write programs to compute them, but because there are more functions than there are programs: there is an uncountably infinite number of functions, and only a countably infinite number of programs — in total. Similarly, there are mathematical states of affairs whose existence can never be proved — not because we aren't smart enough to figure out the proofs, but because there are more states of affairs than there are proofs: there is *at least* an uncountably infinite number of mathematical states of affairs involving just the real numbers, and only a countably infinite number of mathematical proofs — in total. Thus, for example, the fact that the set Z is the union of the set X and Y is a state of affairs. But there is an uncountably infinite number of these states of affairs if X, Y, and Z are each sets of real numbers, and only a countably infinite number of proofs — in total.

Suppose we have an 8-1/2-by-11-inch piece of transparent paper that has been ruled with a small rectangular grid. Suppose we are given some black squares, each the size of a square in the grid, except that the number of black squares is significantly less than the total number of squares on the piece of paper. Now place an 8-1/2-by-11-inch photograph of Lincoln behind the piece of paper. Our goal is to copy, as well as we can, Lincoln's image using an appropriate arrangement of the black squares. Now since the number of black squares is signficantly less than the number of squares on our paper, we cannot hope for arbitrary accuracy everywhere on our paper. We can use all our squares to reproduce as accurately as our grid allows, a portion of the photograph, but the rest of our paper will be blank. Or we can distribute our squares over the entire sheet of paper, and derive a poor reproduction of the entire photograph. Or we can choose some distribution in between.

Suppose, at the quantum level, the problem we face is similar: there simply aren't enough black squares to go around. It has nothing to do with our lack of knowledge as to the "underlying reality". Suppose we say to physicists: "Look, you can't have everything! You can't have infinite mass at every point in the universe. You can't have massive bodies traveling at the speed of light. And you can't have more black squares than Nature provides. Period. End of story."

Another expression of the same idea. Suppose all matter consisted of stones of a certain size and shape. Suppose further that if I wanted to make a model of the mountain I was standing at the base of, I would have to take some of the stones from those that made up the mountain. It seems clear that the smaller the stones, the more accurate could be my model for a given total weight of stones to be used in the model. The bigger the stones, the less accurate. Of course if I were allowed to increase the total weight of stones to be used, then I could make my model more accurate — but I would have to take more of the stones that made up the mountain.

Still another expression of the same idea: suppose that there is an upper limit to the amount of information that can be transmitted through a volume of space per second. This upper limit is a physical limit, and thus is not a function of how clever we are in developing communications devices. Suppose, now, that the amount of information we, at one end of the craps table, would need in order to predict how a pair of dice thrown at the other end, will fall, is simply greater than the upper limit for transmission of information through the volume of space in between. Then we could not make an accurate prediction, and although the reason would be that we didn't know enough, it would also be that it was impossible to know enough.

Waves

"So what is this mind of ours: what are these atoms with consciousness? Last week's potatoes! They now can *remember* what was going on in my mind a year ago — a mind which has long ago been replaced.

"To note that the thing I call my individuality is only a pattern or dance, *that* is what it means when one discovers how long it takes for the atoms of the brain to be replaced by other atoms. The atoms come into my brain, dance a dance, and then go out — there are always new atoms, but always doing the same dance, remembering what the dance was yesterday." — Feynman, Richard, *What Do <u>You</u> Care What Other People Think?*, W. W. Norton & Company, N.Y., 1988, p. 244.

"'The wave is not the water. The water told you about the wave going by. But the wave has a patterned integrity of its own — absolutely weightless.'" — Fuller, Buckminster, quoted in Kenner, Hugh, *Bucky: A Guided Tour of Buckminster Fuller*, William Morrow & Company, Inc. N.Y., 1973, p. 98.

Research project: find out as many applications as you can of the phenomenon — of the *idea* — of "beats", i.e., the fact that when two sound waves of nearly identical frequencies are sounded together, they produce a low-frequency sound, and as the frequencies of the two waves grow farther apart, the beat frequency increases.

"Simple example: the *waa-waa* sound you hear when two instruments play out of tune. more generally, "beats" is just a specific example of interference of waves, and that is literally *everywhere*. Examples: diffraction, e.g. when a wave passes through an aperture — this puts fundamental limits, e.g., on astronomical telescopes. Anti-reflecting coating on glass: the coating reflects an additional wave that interferes with (cancels in this case) the wave reflected off the glass. Etc." — E. B.

Research project: determine the function (if one exists), that, with suitable parameters, allows one to create a wave that moves, at any speed, in either direction, along a horizontal axis.

"Your function is any function of the form $F(x,t) = f(x + vt)$, where x is horizontal position, v is velocity (positive or negative), t is time, and f is *any* function. Physicswise, actual information transfer is bounded by the speed of light." — E. B.

Thus we can consider a sine wave in the x-y plane, and then for each t coordinate, with the t axis running perpendicular to the x-y plane, and with the positive direction into the page, we can imagine the sine wave shifted to the right in proportion to the magnitude of the t coordinate. So:
$F(x, 0) = \sin x$;
$F(x, t) = \sin (x - t)$.

On a windy day, the peaks of the waves moving toward the shore of San Francisco Bay seem more or less the same distance apart. Is there a general relationship between this distance and the velocity of the wind, and if so, what is it?

"Water waves are a very complicated subject, because you're out of the idealized world physicists sometimes like to play with and into the real world. Ignoring wind for a moment (i.e. retreating to the idealized world), *shallow* water waves (shallow relative to wavelength) have velocity that is proportional to the square root of the depth of the water, whereas deep water waves have velocity proportional to the square root of the wavelength. My guess is that this is the dominant determiner of speed, and the wind speed mainly determines amplitude — i.e. when the wind is blowing at the natural wavelength for that patch of water, the wind will reinforce the waves strongly and make them bigger, a sort of resonance effect. But this is just a guess." — E. B.

Why have physicists always been so perplexed by the fact that light has particle properties (the particles of light are photons) as well as wave properties, when they have no trouble with the fact that air and water have both these properties (air and water are each composed of atoms and molecules, but they also exhibit wave properties (sound waves, water waves))?

"Good point: photons behave most like classical waves when there are lots of photons moving together coherently. this sounds pretty similar to the water situation. But there are a couple of key differences:

"*i*) The water is happy just sitting there, and then waves are compressions of density or surface waves on top of this pre-existing medium. Light, on the other hand, is itself a wave. Until around 1900, people thought that there must be something waving with light, so they made up this stuff called the "ether". It turns out that's not there. But if it were, the ether would be more analagous to the water, and the light would be analagous to the water waves.

"*ii*) The photons *always* have both particle-like and wave-like properties. If you turn down the intensity of light enough so that you only let out a single photon at a time, and that photon passes through a grid of slits, you will still get wave-like interference on the other side. A single photon interferes with itself! This is weird. A single photon still has wavelike properties — it doesn't need to be in a herd of photons. The same goes for a single electron or neutron etc. One of my professors actually did the experiment with neutrons: turn down the intensity until only one hits the slits at a time, and it still interferes with itself. Cool!!! A single water molecule, on the other hand, will have no wavelike properties. (Or rather, it would behave like a quantum mechanical wave, but this is a distinct wave, with different properties, from the macroscopic water wave.)

"In general on waves, and more generally than that: I was TA [teaching assistant] this semester for a class for non-physics undergrads. The professor had a vision for what this class should look like, and he has written a book which is now free on the web (eventually to be published). He calls his class 'Physics for Future Presidents', and his goal is to teach the material someone in government would need to interact intelligently with a science advisor and make policy decisions. Global warming, nuclear weapons, aliens, hydrogen economy, lasers, etc. I can't say enough good things about the book. I think 30 years from now people will look on it as a revolution in science education, much like the Feynman lectures are for physics majors. Anyway, chapter 7 is on waves, but the whole book is super. Go to muller.lbl.gov and click through to the course website." — E. B.

Water Draining in the Bathtub

And what *does* make the water go around when it drains in the bathtub or sink or toilet? Get any college physics textbook and work through the formulas for the Coriolis effect — which accounts for the vortex patterns of the weather, including hurricanes — and you will easily convince yourself that the same forces are far too weak to exert such a strong effect over the short distances involved in tubs, sinks, and toilets. Some obvious experiments suggest themselves: (1) Beginning with an empty tub, allow only a very little water to enter the tub and drain, then gradually increase the amount and observe how the flow pattern changes — i.e., run the normal course of events in reverse; (2) Try different sizes of round drain holes, ranging from, say, the diameter of a pencil up to, say, a diameter of two inches; (3) Do the equivalent with square drain holes; (4) Find an expert on fluid mechanics who is willing to talk to you, and discuss the possibility that a circular flow may, in fact, be the most efficient way for the typical quantities of water involved, to flow through a circular opening — for the molecules to "line up".

This subject is discussed in Martin Gardner's *The New Ambidexterous Universe* (W. H. Freeman and Company, N.Y., 1990, pp. 48-51). He concludes, "No one doubts that the Coriolis effect is responsible for the strong tendency of cyclones and tornadoes to spin counterclockwise in the northern half of the globe and to go widdershins in the other half...As for bathtub vortices, the question is still controversial, calling for bigger and better-controlled tubs before any final verdict can be rendered." (p. 51)

"It's a common fallacy, even among some physicists, that the Coriolis force determines bathroom drainage. (The Coriolis force does determine hurricanes' rotation.) Physicists I trust say (and this makes sense to me) that residual angular momentum in the water is what actually does it: since angular momentum is conserved, if there's only a tiny circular flow far away from the drain, it will become a very fast circular flow as it's sucked in toward the drain. This initial tiny flow could be set up any number of ways, e.g. by you getting out of the tub or lifting the plug or ..." — E. B.

What Causes Vortexes?

Vortexes do not merely occur in bathtubs and sinks, or in certain weather patterns, where they can be explained by the movement of air over the rotating earth. We see them also in space, e.g., in the very form that galaxies often take, including our own Milky Way galaxy. Matter streaming into a black hole (as is presumed to exist at the center of our galaxy) seems to follow a vortex pattern. Is there a relatively simple explanation for this phenomenon? Why shouldn't matter simply move straight into the black hole from the point at which it first experiences the gravitational force of the black hole? Why take this roundabout way of getting there?

"Let's talk about matter going into a black hole (the galaxy situation is analagous): What's happening here is essentially an orbit: matter is whizzing by the black hole at high velocity, and the black hole sucks it in. But it only sucks hard enough to bend the trajectory of the matter, because the matter is moving so fast. It keeps on bending the matter towards it, while the matter wants to keep on going in the most straight path available to it. If the gravity is strong enough, the sucking eventually wins out; if it's not strong enough, the matter will fly off into space. This is just what happens when a satellite orbits the earth: the satellite is falling all the time towards earth, but its sideways velocity is just big enough that it always misses the earth as it falls. See chapter 3 of that [online] book I mentioned above. If the matter started from rest, it would just go straight in, no curving." — E. B.

The Distance from the Earth to the Sun

Each planet, including the earth, follows an elliptical orbit around the sun, with the sun located at one focus of the ellipse. Thus the distance from any planet to the sun is continuously varying. So why are children in primary school and secondary school taught that the "distance to the sun from the earth is 93 million miles"?

Why Do Planets Travel in Elliptical, as Opposed to Circular, Orbits?

I have never come across an explanation, in undergraduate textbooks, or in popularizations, as to why planets travel in elliptical orbits. The kind of explanation I have in mind would begin with a planet-sized object at some distance from the sun. Then among the possibilities are: (1) the planet is moving directly toward the sun; (2) the planet is moving continually away from the sun; (3) the planet is at a fixed distance from the center of the sun, and not moving relative to the surface of the sun; (4) the planet is moving such that, at any given moment, a component of the vector of its movement is tangent to the surface of the sun.

Regarding case (4): we know from high-school physics that if we throw a ball in a direction parallel to the surface of the earth, then if we throw the ball sufficiently fast, and ignoring air resistance, it will go into a circular orbit around the earth. In effect, the earth keeps "dropping away" at a rate equal to the falling rate of the ball. So if the planet is moving similarly, it should be in a circular orbit. Why isn't it? If the tangential speed of the planet is greater than that required to make it go into a circular orbit, does it go into an elliptical orbit? If so, can we attribute the elliptical orbits of planets to the fact that there are more ways for the tangential speed *not* to be the exact speed required for a circular orbit?

Finally, is there a way to prove, by considering only straight-line approximations to the movement of the planet[1], and straight-line approximations of the speed-of-light travel of the particles of gravity, and Newton's well-known equation for the force of gravity, $F = (Gm_1m_2)/r^2$, where m_1 is the mass of the sun and m_2 is the mass of the planet, to show that the orbit of the planet must be elliptical?

Eratosthenes' Calculation of the Circumference of the Earth

Most histories of science devote a few words to the ingenious calculation, by Eratosthenes (c. 273 - c. 192) of the circumference of the eath. The histories describe his method as follows:

When the sun at the summer solstice is directly overhead in Alexandria, it casts a shadow from a stick planted vertically in the ground some 500 miles away, at Syene (modern Aswan). From the angle that the far end of the shadow makes with the vertical stick, one requires only some basic geometry to determine the circumference of the earth.

The problem is that none of the histories of science that I have seen explain two crucial things. (1) How is it possible to know at Syene, when it is 12 noon on the day of the summer solstice at Alexandria? There were no electronic means to communicate from Alexandria, "It is now 12 noon here," and there were no clocks. Even if there were, they would have to be able to keep time while being moved from Alexandria to Syene. (It was not until the late 1700s that a clock was developed, by John Harrison, that could keep time while in motion, in this case, aboard a ship on the high seas.)

The answer to question (1) is that it will be high noon in Alexandria when the shadow thrown by the stick in Syene, is the shortest[1]. The observer can arrive at the shortest length by pressing a peg, or moving a white stone, at the end of the shadow throughout the morning.

The second question, (2), is, How was Eratosthenese able to measure the small angle (around 7.2 degrees) defined, at the top of the stick, by the length of the shortest shadow? It seems highly unlikely that any instrument available to him was capable of measuring 7.2 degrees. One answer is the following:

He knew the length of the vertical rod. Call it a. He could measure the length of the shadow at noon. Call it b. Then by the Pythagorean Theorem, he could compute the hypotenuse of the right triangle so formed, that is, he could compute $h = \sqrt{(a^2 + b^2)}$.

Now, on a sufficiently large, flat, area of fine-grained sand, he could draw a circle of radius h. Then he could draw one actual radius in the sand.

He could then measure the length a of the shadow on this radius, with the one end of the length at the center of the circle. At the other end, he could erect a perpendicular at the end of b on the radius (it is reasonable to assume that wooden right triangles were available) and extend it to the circle.at a point p. He could then exend the first radius until it met the circle at a point q.

He could now measure, using a rope, the arc lying running from q to p. Call this length r. Then the circumfernce C of the earth can be found from the equation,

$$(500)/C = r/(2\pi h)$$

By this method, Eratosthenes arrived at a value of C that is only a couple of hundred miles from what is now considered to be the actual circumference of the earth, namely, 24,800 miles.

---

1. where the vector of this movement must have a non-zero component parallel to the surface of the sun
1. Stephen Strazdus, in a note on the web site titled "How did Eratosthenese know the exact, accurate hour to measure the difference in sun angle (7 degrees) at 2 cities at the very same time?", https://www.quora.com (5/30/17).

The Three-Body Problem

I have never seen explained in any discussion of the three-body- (much less the *n*-body-) problem, why a solution (in this computer age) is not simply to take a typical calculus approach: i.e., break time, distances and gravitational attraction down into finite units, then make these as small as needs require. We begin with each body in some initial position and moving at some initial velocity. We now divide the distances between all the bodies into finite intervals. In the first time interval, starting at time t_0, each body emits its appropriate "gravitational attraction pulse" and moves the appropriate number of distance units in the direction prescribed by the body's current velocity. The gravitational attraction pulse travels the number of distance units prescribed by the speed of light. The process repeats starting at the second time interval, t_1. (How is the *n*-body-problem solved in video space-war games?)

"You're correct, people solve the 3- and *n*-body problems numerically all the time. This is a lot of what working planetary physicists do. As a practical matter, this is entirely sufficient — you can get as much accuracy as you desire by throwing more and more computing power at it. But aesthetically, it bugs people that there is no analytical solution to such a simple problem." — E. B.

Note: The chemist Louis Whaley has informed me, "There is a solution to the 3-body problem in Quantum Mechanics; the hydrogen molecule (2 protons and 1 electron) can be solved analytically in elliptical coordinates."

Air Flow Over a Wing

The standard explanation of why air flowing over a wing produces lift involves the Bernoulli effect: since the distance the air must travel from the front of the wing to the back is less when the air flows along the lower surface than when it flows along the upper, the speed of the air along the upper surface must therefore be greater than the speed along the lower, and this produces a lower pressure on the upper surface than on the lower, resulting in lift.

But this explanation assumes that two air molecules at the leading edge of the wing, one of which will pass along the upper surface, and the other of which will pass along the lower, are somehow "obligated" to reach the trailing edge of the wing at the same time. Why should this be so? Why doesn't the upper surface molecule travel over the surface at exactly the same speed as that of the lower surface molecule, and simply arrive at the trailing edge at a later time?

"Your point might be valid for any given *pair* of molecules, but if you let this happen in aggregate, you'd get discontinuities in the airflow. That is, on average not as much air would go along the top, so you'd get regions of low pressure up there (vacua), and the pressure would accelerate more air in there to make up for it. That's a long way of saying that your answer lies in the continuity equation.

"By the way, my fluid mechanics book says that wing angle — the fact that the wing is tilted so that the bottom hits the oncoming air and gets bumped up by it — is as, or more, important than the Bernoulli effect. In practice, you solve air-flow equations which have both effects embedded in them in complicated ways." — E. B.

Probability and Determinism

Suppose I am standing at one end of a craps table, and I press two fair dice between my thumb and index finger, with, say, the 1 and 6 faces pointing up. Suppose I now slowly move my hand, holding the dice with these two faces point up, to the other end of the playing surface of the table. While I am doing this, there will be no doubt in the minds of onlookers that, if I continue to hold the dice in the same position, when I place them on the playing surface, the 1 and the 6 will be point up.

Now suppose instead that I shake the dice in my hand, then send them flying in the traditional way over the playing surface. Onlookers will accept that they will not know which two faces will be pointing up until the dice stop moving. They will say that the best they can do, as far as predicting which two faces will be pointing up, is to state the probabilities of each possible pair of faces poing up. Thus, e.g., the probability of a 1 and a 6 is 2/36.

One or more of the onlookers may remark that it is not possible to know in advance which two faces will point up. Someone may remark on non-determinism: there are some things we cannot know exactly in advance.

But suppose someone replies that determinism has nothing to do with it. The problem is that the information highway between the dice in mid-flight and the observer is simply not able to convey all the information needed to enable an exact prediction of which two faces will point up.

This problem did not exist when I held the dice pinched between my thumb and index finger. But when I threw the dice, I would have needed to know the exact mass of each die, its exact orientation and position above the table at the moment of release from my hand, and then been able to compute the position of each die as it flew through the air. This would have required considerable computation and measurement — the information highway between the dice at any point and my eyes, would have been overloaded.

Or suppose that we are receiving a message in Morse code transmitted over a wire. But there is noise in the wire, and so some of the dots and dashes are missing or garbled. We might argue that this is because the message is inherently not deterministic — there is no unique message, only a probabilistic message. Or we might argue that the transmission line is faulty.

Or suppose we are looking at a person who is physically far away from us. We might argue that the features of the person are inherently not deterministic — there is not a unique person, only a probabilistic person.

So perhaps we can argue that the universe is deterministic, always, but the information highway between events and observers is sometimes overcrowded.

Negative Probability

Project: write a report, for the mathematically-educated layman, on the subject of negative probability, and why such a thing cannot exist.

"Eine genauere Analyse dieses mathematischen Schemas hat es aber wahrscheinlich gemacht, dass jene Grössen, die in der gewöhnlichen Quantentheorie als Wahrscheinlichkeiten gedeutet werden müssen, hier unter Umständen negativ werden können, nachdem der Renormierungsprozess durchgeführt ist. Das würde natürlich eine widerspruchsfreie Deutung des Formalismus zur Beschreibung der Materie ausschliessen, denn eine negative Wahrscheinlichkeit ist ein sinnloser Begriff."

"A more precise analysis of this mathematical approach [to combining relativity and quantum mechanics] would probably reveal that every quantity that must be represented, in normal quantum theory, as a probability, could, in some circumstances, become negative, because of the renormalization process. That would of course exclude a contradiction-free interpretation of the formalism for the description of matter, since a negative probability is a nonsensical concept." — Heisenberg, Werner, *Physik und Philosophie*, Ullstein Bücher, Berlin, 1978, p. 135. (Translation mine.)

"If you want probability to mean the number of times some event will happen out of a given number of test cases, that number better be positive to make any sense. That is, if you flip a coin a hundred times, the number of times you get heads better be a number between 0 and 100. That's it. Negative numbers just don't make sense in this context. We take it as a basic consistency condition of our quantum mechanical theories that they yield sensible numbers for probabilities (not negative, not infinite — restricting to non-infinite probabilities is one of the major constraints we have on string theory, actually.)" — E. B.

The Birds-in-the-Truck Problem

Project: write a review of the status of the Birds-in-the-Truck Problem. The Problem is as follows: a man is driving a truck full of birds through the countryside. He comes to a bridge which can only hold the weight of the truck without the birds. What should he do? Common answer: he should scare the birds, e.g., by clapping his hands, so that they fly up from their perches, hence remove their weight from the truck, so that he can then drive across the bridge. Common counter-argument: the air-presssure from their flapping wings would still increase the weight of the truck.

"I suppose it's true that momentarily, the birds must exert a force greater than their weight down on the truck or on the air around the truck to lift off, but this would only be temporary. Maybe it would be easier to think of monkeys which had trampolines they could jump off of." — E. B.

Math-in-Physics Anxiety

"In Figure 3-19a we see an elevator at rest. Let us consider an object such as a bag hanging from a spring scale as shown. The scale indicates the downward force exerted on it by the bag. The force, exerted *on* the scale, is just equal and opposite to the force exerted by the scale upward on the bag. We call this force, T. Since the mass, m, is not accelerating, we apply $F = ma$ to the bag and obtain

$$T - w = 0,$$

where w is the weight of the bag; thus $T = w$, and since the scale indicates the force T exerted on it by the bag, it registers a force equal to the weight of the bag, as we expect. Now, if the elevator has an acceleration a, when we apply $F = ma$ to the object, we obtain $T - w = ma$. We then solve for T and obtain

$$T = w + ma."$$

— Giancoli, Douglas C., *Physics*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1980, p. 75.

I certainly understand that the equation, $T - w = ma$, means that if I subtract the weight of the bag from the force exerted on the scale by the bag, I will be left with the quantity given by multiplying the mass by the acceleration. But when the text says, "We solve for T and obtain $T = w + ma$," that is by no means a worry-free statement for me *in physics*! Of course, I know that, according to the rules of algebra, I am allowed to move a term from one side of an equation to another as long as I change its sign. But why should I be sure that I can move the *weight* to the other side by changing *its* sign? How do we know we are dealing with real numbers that follow the rules of mathematical real numbers when we measure physical quantities? The answer, "Because that is how we define physical quantities" does not seem satisfactory. If I define certain emotional states to be quantities that obey *some* of the laws of equations (e.g., the law concerning changing of signs when terms are moved across the equals sign), and if I then write, e.g., "hope = evidence + past experience", you would probably object if I concluded, "and therefore past experience = hope – evidence."

"What does one mean by a force? something with an amount (of pushing) and a direction (of pushing). It is thus very naturally a vector. We have algebraic rules for how to manipulate vectors, and they make sense too: $T - w = ma$ says: the acceleration (times the mass) is given by the sum of the forces on the object. $T = w + ma$ says: to produce a given acceleration, given that the block has a certain weight, what tension do I need to apply?

"Note that because force is very naturally a vector, adding and subtracting is also very natural. This is very different to the situation you have with 'hope' and 'evidence', which are not in any natural way vectors or even numbers.

"But I am sympathetic in general to your concerns. I wondered for a long time why Newton's equation $F = ma$ is a second-order differential equation in position. Why not third or higher order? Why is it that the fundamental equations of motion are second order? I can wave my hands now and talk about how the equations come from approximations to the fundamental equations of quantum mechanics, namely extremization problems on path integrals, but it's still not totally satisfactory. At some point, you have to retreat as follows: the job of physics is to give equations that can predict the results of experiments. The 'why' of 'why do things work like this?' will probably never be completely answered (though we try to do better and better)." — E. B.

Custom Graph Paper

Why don't we plot at least some non-linear functions on the appropriate analogue of log graph paper? Why don't we have "acceleration paper" to plot the speed of falling (or rising) bodies? If we did have, ready-to-hand, such graph paper for any non-linear function, would that make the "problem" of non-linearity largely disappear, and if not, why not? Why shouldn't everything always be as linear as it can be?

"You could, of course, do this, but we don't because it's easier (most seem to think) to look at lots of different graphs on the same graph paper than lots of the same graphs on different graph papers. I don't know — the opposite may be true for you." — E. B.

Why Don't Formulas Have Addition Where They Have Multiplication?

Why are the formulas of physics as they are insofar as operations are concerned? For example, why is momentum equal to mass times velocity instead of mass *plus* velocity? And similarly for force (mass times acceleration), work (force times distance), etc. The answer sought here is not "Because that is the way these quantities are defined," or "Because that is what experiment reveals," but rather something like the explanation why the force of gravity is an inverse square law, namely, because if we think of the gravitational force as radiating from a point mass and being measured by the intensity per unit area on successive concentric spheres about the point, then in fact that intensity turns out to follow an inverse square law, by definition of the area of a sphere.

"First, regarding gravity: there is no *a priori* reason gravity had to be inverse-square. It's convenient and cute that there's a relation to how intensities per unit area fall off, but it didn't have to be that way. The strong nuclear force actually gets stronger, linearly, with increasing distance (at least at some energy scales — the behavior changes with energy). The weak nuclear force falls off exponentially with distance. Different electric and magnetic configurations have forces which fall off with various powers of distance (though the fundamental electrostatic interaction between point charges is inverse-square). Furthermore, gravity is only a good approximation to inverse-square, and there are general relativistic corrections which depend differently on distance.

"I think it's wrong to discount 'because that is what experiment reveals', because the fundamental goal of physics is to describe mathematically what happens in an experiment, however you get it done.

"We can still make some sense out of things. Let's think about momentum. Momentum is just force applied for an amount of time. But $F = ma$, and acceleration for a time gives a change in velocity, so naturally change in momentum is mass times change in velocity. Also, the units work out.

"Then why $F = ma$? Because Newton said so. But even then, this makes sense: say you fix the mass you're talking about. To accelerate it twice as much, you need to push twice as hard, right? So the force better be proportional to acceleration, with some constant of proportionality, k: $F = ka$. Well, you do some measurements, and you figure out that all k depends on is the mass." — E. B.

How Do You Discover Something "Is Proportional To" Something Else?

How did physicists in the past decide that one quantity was "proportional to" another? At what point in a series of experiments do I decide that something is proportional, or inversely proportional, to something else? Don't I first see the actual ratio, with the constant of proportionality already in place?

"You can determine proportionality without determining the constant. Example: your average spring applies a restoring force to a mass on its end which is proportional to its length: $F = kl$, l being length. How would you determine this? You might hang the spring from the ceiling and start hanging different weights from the spring. you hang a bowling ball from the spring and you measure how much it stretches. You hang 2 bowling balls and note that it stretches twice as much. You hang 3 bowling balls and note that it stretches three times as much. Now if you know how much each bowling ball weighs, you can determine k. If you don't know how much each

bowling ball weighs, you can't determine k. But you have learned that force is proportional to *l*."
— E. B.

"Take the Limit as the Size of Electrons Approaches That of Photons..."

Compare Young's double-slit experiment with the one that yields the fundamental dilemma of quantum mechanics. The only difference is that the particles are photons in the one, electrons in the other. But electrons are bigger than photons. Do we gain any insights if we start with Young's experiment, and then gradually (continuously) deform it into the quantum mechanics experiment?

"I don't know what you mean by saying 'electrons are bigger than photons'. In the standard model of particle physics (including QED [quantum electrodynamics]), both the photon and the electron are point particles. Sometimes people talk about a 'charge radius' of the electron, which means the following: the quantum-mechanical vacuum surrounding the electron is actually a very active place, and in some sense there are electron-positron pairs appearing and then disappearing all the time. The electron attracts the positrons, and so there is a cloud of virtual positrons around the electron. This reduces the electric field far outside the electron, the positively charged positrons cancelling some of the field of the negatively charged electron. If you shoot something at high speed toward the electron, it can eventually start penetrating the positron cloud and seeing the 'bare charge' of the electron. The rough distance from the electron at which this effect becomes important is called the 'charge radius' of the electron. But this wouldn't affect the double-slit experiment.

"All that really matters for the double-slit experiment is that both photons and electrons behave like waves. The amount of diffraction caused by the slits depends on the size and spacing of the slits relative to the wavelength of the electrons or photons. The wavelenths of the electrons and photons, in turn, depend on their momentum. You see the same diffraction effect with plain old water waves." — E. B.

Particle Physics, or, the Furniture Maker

Suppose that extraterrestrials conduct a series of experiments to determine what types of furniture exist on earth. For example, they have an earthling bring plans for a type of chair, call it chair type X, that the extraterrestrials haven't observed before, to a furniture maker. Within a month, the earthling returns to the shop and picks up the chair. The extraterrestrials conclude, "Chairs of type X exist!" They repeat the experiment with many different types of chairs, tables, cabinets. Each time, they add another furniture type to their growing list of types of furniture that exist.

How is this process fundamentally different from physicists' search for new particles? What does it mean to perform an experiment, observe a particle that has never been observed before, and conclude that the particle exists? What is the difference between saying "Particles of type X exist" and "Particles of type X can be created?"

Heat: What Makes the Particles Move in the First Place?

We all know that heat is the total kinetic energy of particles — of a liquid or a gas — in some region, e.g., in a balloon. And kinetic energy is a function of the speed and mass of the particles. The question is: what gets the particles moving in the first place? Consider a large empty space (a vacuum) surrounded by metal. Inside there is exactly one gas particle. Assume it is perfectly still relative to the sides of the container, and located in the middle of the container. Now heat the container. What exactly makes the particle start to move? We assume that no particles are given off by the interior of the container itself.

"[The answer is] radiation. Anything (at least anything made up of charged particles, meaning just about anything but neutrons) heated above absolute zero radiates electromagnetic radiation (i.e. light). Electromagnetic radiation is created by jiggling charges. You and I are made up of jiggling electric charges, and we radiate light. We radiate at many frequencies, but most of the radiation comes out in the infrared, which is too low in frequency for our eyes to detect. That's what passive infrared goggles pick up: the infrared radiation coming off our bodies. Same thing with the box surrounding your atom. It will be at some non-zero temperature, and so it will radiate. The atom will absorb some of the radiation, turning the radiative energy into kinetic energy. Insterstellar and even intergalactic space are not empty: There are protons, neutrinos, and photons running around (and other stuff). Everywhere there is at least the radiation left over from the big bang. So no matter what, you always have hot things around, hot things radiate, and so it's impossible to perfectly isolate any particle.

"That said, physicists can now do a pretty good job: they cool things with liquid nitrogen (to about 70 K), then inside they have another box cooled by liquid helium (about 4K), and then inside that they have nice vacua and they let things expand quickly, which cools them, etc. and they can hold things at millikelvin temperatures. But it's impossible to get to absolute zero." — E. B.

Is Size a Place?

If we had the ability to shrink things, then we could hide something by making it sufficiently small. Or, if there weren't enough room for something we wanted, e.g., a big house, we could just shrink it (and ourselves). We could solve the problem of the homeless by simply shrinking them at night.

Why Not Just Combine Quantum Mechanics and General Relativity and Call the Result "The Theory of Everything?"

"[It is true that] we can explain every experiment we have done to this point using QM [Quantum Mechanics] or GR [General Relativity]. (There is one exception to this which I'll mention below.)

"1) Quantum mechanics is so basic to our understanding of physics and comes from such basic principles, that we expect it to be relevant to GR. Quantum mechanics starts from the observation that "measurement is disturbing." If I measure you, say, by collecting a bunch of photons bouncing off you into my eye, you are not bothered much, because you are a big macroscopic thing made of many electrons and other particles, and I don't have detailed experimental control over what each is doing. However, if I try to measure a single electron by bouncing a photon off of it, the photon will disturb what it's doing, and I can have enough experimental control over the electron to notice this. One of the consequences is Heisenberg's uncertainty principle: You can't simultaneously measure with perfect accuracy the position and velocity of the electron. Since it stems from such a simple (and in retrospect clear) idea that "measurement is disturbing", we expect similar principles to apply in gravitational situations.

"2) So, let's try to make GR a quantum mechanical theory. Turns out that it's not so easy to do this. The quantum mechanical particle that would carry gravitational force in a quantum theory of gravity is called the graviton. If you compute the probability for two gravitons to scatter off one another in the most naive formulation of quantum gravity, the probability turns out to be infinite. This is bad, clearly. The naive theory is failing.

It's not so easy to explain intuitively why this is happening, but it's something like this: one of the consequences of QM is that a particle never has perfectly-well-defined energy. And in particular, it has some (very small) probability of having arbitrarily high energy. But the strength of the gravitational force depends on the energy (which, as Einstein taught us, is the same thing as mass). So arbitrarily large forces appear in your problem, and even at very suppressed probabilities, these muck things up for you.

"3) So one thing you could try to do is to explicitly cut off the range of energies you allow in your problem. We do this all the time in quantum field theory (the version of quantum mechanics that we use now), and we can do it rigorously (following the procedure of Wilson): you start with a theory valid at arbitrarily high energies, and there is a procedure (Wilson's) for taking account of all the high-energy effects and arriving at a different low-energy theory with a cut-off range of energies. The two are equivalent but different descriptions of the low-energy phenomena.

"This is what we do, essentially, when we do chemistry: we use the low-energy theory of molecular interactions, already having taken account of the high-energy nuclear processes.

"4) The problem with this approach is that if you want to do nuclear physics, say, your effective molecular theory isn't worth much. If you don't know it, you need to "guess" the nuclear theory which "completes" the low-energy picture.

"This is analagous to where we are now with QM and GR. We do have a perfectly good quantum theory of GR, but it is an effective one, with an explicit cutoff in energy ranges. This is theoretically unsatisfactory because we know there must be some more fundamental theory valid at all energies, and it may eventually be experimentally unsatisfactory if we can do experiments at much higher energies than we can do now (say, near a small black hole).

"5) In summary, for this part of it, the underlying principles of QM are so simple that we expect them to be applicable to gravity, and we have a consistent quantum gravity theory, but it is only valid at low energies. We know there is something beyond it. This something may be string theory or something else yet to be discovered. As of yet, the possible somethings (including string theory) are only poorly understood, and none is yet a convincing candidate.

"6) It turns out that there are also good theoretical reasons to expect that our current quantum field theory of the nuclear and electromagnetic forces (the 'Standard Model') is also only an effective theory. Theorists expect (hope?) to see signs of the new theory at the LHC collider in Geneva when it turns on in 2007.

"7) There is one very disturbing problem, an experimental problem really, that is still completely not understood within our QM + GR framework: We have learned recently that there is a small, non-zero cosmological constant. This is the "dark energy" or "vacuum energy" or "antigravity" which is making the expansion of the universe accelerate. No one has even the beginnings of a good idea of what this stuff is or why there is as much of it as there is.

"8) There are properties of the 'Standard Model' we would like to understand better. Why is there such a huge range of masses for the various particles? (The top quark is 10^5 to 10^6 times heavier than the electron, not to mention neutrinos which are probably much lighter.) Why is the mass of the electron what it is anyway? Also, the different sorts of particles seem to arrange themselves into a 'periodic table'. In chemistry the periodic table was a big hint to some underlying structure (which was discovered to be atomic structure in the early 20th century). What structure underlies the standard model?

"It may be that things just are the way they are, and we'll never understand why the electron mass is what it is. But physicists want to understand, and for now they'll keep on trying." — E. B.

Is a "Theory of Everything" Impossible Because of the Ever-Increasing Sophistication of Scientifc Instruments?

In the 1930s, Gödel proved that in each mathematical system sufficiently robust to contain basic arithmetic, there are truths that are impossible to prove. Furthermore, if you add these truths as axioms, there will always be other truths that are impossible to prove.

Algorithmic information theory teaches us that some finite binary strings have short minimal representations, others longer ones, others have minimal representations that are as long as the strings themselves (these are called "random" strings of binary digits). A short minimal string can be regarded as a "theory" of the string it represents.)

A naive view of physics is that there is a fixed set of physical entities — mass, energy, momentum, wavelength, velocity, acceleration, time, distance, ... — and that physicists try to discover better and better descriptions of the relationships between these entities, while the makers of scientific instruments (including of telescopes), keep improving their products so that physicists can obtain more and more accurate data about this world.

A more realistic view is that improvements in scientific instruments — everything from particle colliders to telescopes, and not only those that view visible light — reveal phenomena that no one had thought of before. Current theories must therefore be modified to take into account these phenomena.

"Almost every principle [in physics] once proclaimed has been subsequently superseded. No matter how useful they are or how good an approximation they give to phenomena, sooner or later most principles fail, as experiment probes the natural world more accurately." — Smolin, Lee, *The Trouble With Physics*, Houghton Mifflin Company, N.Y., 2006, p. 218.

Thus it is hard to believe, at least for me, that there ever will be a single, permanent "Theory of Everything" in physics, because each theory in physics is ultimately based on what the technology of the time is able to measure. The only way for a theory to be "permanent" would be to stop development of all measurement technology beyond that which seems to confirm the theory.

Since many, perhaps most, physicists must long ago have thought what Smolin expresses above, I have to ask myself why a Theory of Everything continues to be held aloft as a goal of physics research. In my more cynical moments, I sometimes think it is little more than a way of securing funding.

Is Physics Ultimately a Function of the Number of Physicists Alive to Work On It?

In each age, it is instructive to ask the question, "How many physicists are necessary to understand all of modern physics?" During Newton's lifetime, at least in the late 1600s, the answer may well have been: *one* — or at least less than, say, ten. In the early 21st century, when worldwide there are tens of thousands of physicists, it is hard to give a number. But we must ask to what degree the complexity of modern physics is a function of the number of physicists who are working on physical theories. What would happen to modern physics, if, for example, 90% of physicists suddenly disappeared?

Is the nature of the physical universe ultimately a function of the sophistication of available research instruments, and the number of living physicists? In other words, is the question "What is the nature of the physical universe?" an incomplete question, the complete one being "What is the nature of the physical universe in such-and-such an age?"?

It might be that mathematicians of Newton's time could understand, and develop, mathematical results that were discovered only in the 19th or 20th centuries, and it might be that Newton could understand something of general relativity. But it is hard to imagine what progress he could make without being able to participate in modern physics culture — without having access to the experimental results and thousands of physicists that constitute modern physics.

"What Is the Nature of the Universe?" "What Universe?"

It makes sense to ask, "What were Newton's thoughts about gravity?" It does not make sense to ask, "What were Newton's thoughts about hadrons?"[1] A modern physicist who sets out to find out more about "the universe" is setting out to find more about a universe that contains, among many other things, the many sub-atomic particles. This was most certainly not true of a physicist in Newton's time.

"What Time Is It?" "Is What?"

Numerous popularizations have attempted to explain the theory of Special Relativity to the educated layman. These popularizations are typically written in the plainest language the subject allows, and contain examples, so that the reader feels that he should really be able to understand the theory, at least at the level at which it is being presented. But more often than the authors of these popularizations realize, the reader comes away feeling that, despite the plain language and the examples, something is still eluding him.

I would venture to say that the concept of an "inertial frame" is probably readily understood by most readers: it can be (and often is) represented by, say, a railroad car which is either station-

1. "A composite particle made of quarks held together by the strong force (as atoms and molecules are held together by the electromagnetic force)." — "Hadron", Wikipedia, 1/31/13

ary, or else is moving at constant speed down a straight track. In either case, it is at least plausible to the reader that physical measurements made within the car must yield the same results, regardless whether the car is stationary or moving.

So far so good.

The constancy of the speed of light is likewise readily understood, although the reader may find it difficult to conceive how this speed remains unchanged regardless how fast a light source is moving relative to an observer.

The typical reader can probably also understand, via an explanation such as the one given in the previous sub-section, why it is that clocks in a given inertial frame slow down as the speed of the frame approaches the speed of light

So far so good.

However from here on, at least in my experience, and I suspect in the experience of many readers, things become much less clear. For one thing, I have never seen a clear, informal explanation why distances shrink as the speed of the inertial frame approaches the speed of light. Statements such as

"Consider: the length measurement of a moving object always depends on making simultaneous position measurements of two points (if you want the length of a moving bus, it behooves you to measure the position of the front and the back at the same time)."[1]

seem to be, if not wrong, then at least confusing. I can easily measure the length of a long, winding road in my city by noting the odometer reading in my car at the start, then driving the length of the road, and noting the odometer reading at the end.

If the reader replies that the author of the above passage meant that the moving object is in an inertial frame different from the one from which the measurement is made, I again make the same argument, for we can suppose that the moving inertial frame is in a very long railroad car that moves in parallel past our own railroad car, which is at rest on a track only a yard or two from the track bearing the moving car. We start our stopwatch at the moment the beginning of the object passes us, and we stop our stopwatch at the moment the end of the object passes us. From the difference in times, and knowledge of how fast the railroad car was moving, we can compute the distance of the object. It is not at all necessary that we measure the beginning and the end of the object at the same time.

Even in excellent historical works like Galison's (see above footnote), it is not clear (at least not to this reader) what *tasks* physicists and engineers in the late 19th, early 20th century, were trying to perform in connection with time. It seems that the tasks included:

(A) Set all clocks in some domain to register the same time as a central clock.

(B) From a reading of a local clock, determine the reading, at that moment, on a central clock.

(C) Determine if two events that appear to occur at the same time (in one inertial frame), really did occur at the same time (in another inertial frame), and if not, by how much the times of occurrence differed, which event occurred first, etc.

1. Galison, Peter, *Einstein's Clocks, Poincaré's Maps*, W. W. Norton & Company, N.Y., 2003, p. 22.

(D) Determine what time it is at some other local clock.

Now if, contrary to fact, the speed of light were infinite, then all these tasks would be easy to perform.

For task (A), we could simply set up wire or radio communication between each clock and the central clock, so that all clocks were controlled by the central clock. Throughout the universe, there would be one universal time (Newton's universe).

For task (B), the reading of the local clock would be the reading of the central clock.

For task (C), if two events, e.g., the sudden appearance of two stars, *seemed* to be simultaneous, they would in fact *be* simultaneous.

For task (D), we would need simply to send a message to the other local clock, asking what time it was, and then get the answer back instantaneously. Or all local clocks could be linked to each other.

However, the fact that the speed of light is finite, not infinite, makes the tasks more difficult to accomplish.

So let us see if we can think through some ways to accomplish each task under this limitation.

Tasks (A) and (B) One way we could accomplish these tasks would be to have people bring their clocks to the central clock, synchronize them there with the central clock, and then carry them (vessels containing the sacred fire), back to their local domains. Then throughout the universe, after everyone had returned to his or her local domain, everyone would know the time at the central clock, assuming no clocks ever failed.

If it were possible for light or an electromagnetic wave to carry a message, "This message has been traveling for x seconds," then the time keepers would not have to travel to the central clock, but instead could simply set their local clocks to the time they received from the central clock, plus x seconds.

Tasks (C) and (D) can be performed by using the Lorentz transformation.

These tasks, and possibly others, along with our naive attempts to find ways to accomplish them, at least give us an orientation from which to confront the subtleties presented in books like Galison's.

"In the absence of gravity, it is possible to assemble a collection of synchronized clocks everywhere in the universe. If everyone knows how far they are from the central clock, the central clock just sends out periodically a signal saying "I, the signal, was sent at time T," and then by knowing their distance, everyone knows how much later than time T it is. If you're speeding relative to the army of synchronized clocks, however, you see the distance contracted between you and the central clock, and hence you determine that it's not as late past time T as your stationary counterpart does. Gravity changes this scenario in an interesting way — the extent to which you can arrange an army of synchronized clocks is an important property of a given spacetime." — E. B.

Communicating the Meaning of Left and Right

It is sometimes said there is no way of communicating the meaning of left and right to someone who does not have the appropriate points of orientation.

We might be inclined to argue that if the other person has a video screen, then we can transmit the meaning. The trouble is, we do not know what direction will be *up* relative to his screen. If we send him an image with "This is the upper left corner", he may view the screen so that the message is in the wrong corner.

But consider:

Suppose we have established communication with an extraterrestrial, and that we have somehow established a mutually intelligible language with which we send written messages. He has let us know that he is viewing our messages on a rectangular, non-square, screen.

We now give him the following instructions:

1. Place your communications equipment, including your viewing screen, plus yourself and all necessary life-support equipment, in a rectangular box in empty space. Let the screen be oriented so that its sides are parallel to the sides of the box.
2. Accelerate the box in a straight line perpendicular to one of the sides.
3. Define the direction *up* to be that of an arrow pointing perpendicularly away from the side toward which free objects move in the accelerating box.
4. Now send a screen image with an arrow pointing up, and with "This is the upper left corner" marked in the appropriate corner.
5. Tell the extraterrestrial to orient the image so that the arrow is pointing in the same direction as the *up* defined by the acceleration of his box. Then tell him that the upper left corner is as indicated, that the upper right corner is at the other end of the screen, etc.

Teaching the Concept of Density

Density is normally defined as mass per unit volume. At any given location, e.g., sea level on the planet Earth, this can be converted to weight per unit volume. But I wonder if young students would not find it much easier to understand the concept if it were expressed in terms of its inverse: volume per unit mass (or weight). Each student could be given, say, a one-pound weight in the form of a cube, and then told how big a cube of each other material — aluminum, lead, gold, wood, glass, ... — would be required in order to make the same weight.

Scale Models of the Atom and of the Light-Year

It is a disgrace that you can get a bachelor's degree in the humanities (and probably in many technical subjects) from a major university in the U.S. and not have the slightest idea of the distances and sizes of particles in an atom, or of how far a light-year is. By an "idea" here I mean, via an easily-visualized scale model. So, as a public service, I offer the following:

"If the electron cloud of the iron atom ^{56}Fe [iron] were expanded to the size of a football field, the nucleus would be represented by a pea-sized ball 4 mm in diameter, weighing 6 million tons, and the electrons would be represented by 26 mosquitoes weighing 120 tons each, flying around the pea at distances ranging up to 50 meters." — Franklin Miller, Jr., *College Physics*, Harcourt Brace Jovanovich, Inc., N.Y., 1977, p. 21.

A light-year is a little less than 6 trillion miles. Alpha Centauri, the nearest star to us, is about 4.25 light years, or about 26 trillion miles, away. The solar system has a radius of about 4 billion miles. So let the solar system be represented by a ball having a radius of 1 foot — in other words, let the solar system be contained in a large beach ball. Then, relatively, Alpha Centauri would be 6,500 feet or about 1.2 miles away.

A physicist's view of the matter:

"I just remember order of magnitudes in meters. So, in meters, the size of the nucleus is 10^{-15}, the size of the atom is 10^{-10}, the size of the sun is 10^{9}, the earth-sun distance is 10^{11}, and ... a light-year is 10^{16}.

"Instead of a scale model [of the atom], if I need intution, I might think of meter sticks — e.g. the size of the nucleus [relative] to the size of the atom is as 1 cm is to 1 km." — E.B.

An astronomer's view:

"Living on the surface of the Earth as we do, it is really difficult to fathom how big the Universe is, even our local part of the Universe known as our Solar System (the sun and its immediate surroundings). Oh sure, we can measure these distances and assign numbers to them — the moon orbits some 240,000 miles (384,000 kilometers) from Earth, and the Earth averages some 93 million miles (150 million kilometers) away from the sun. But what are such distances like on a scale we might be able to understand?

"If we squished the Earth down to the size of a good-sized orange or a baseball (about 3 inches in diameter), and scaled everything else down by the same factor, how big would the moon be and how far away from our 'orange' Earth? Well, the moon is a little less than a third the diameter of the Earth, so something like a ping pong ball would be about right. And as for distance away, would it be a few inches? A foot? If you calculate it, the ping pong ball orbits about 6.7 feet (just about 2 meters) away from the orange!

"Now let's compare this to a typical space shuttle orbit. The space shuttle, of course, is unable to go to the moon, even though it does go into "outer space." However, you might be surprised to find out that the space shuttle only ventures out to a couple of millimeters (less than a tenth of an inch!) up off the surface of the orange! On this same scale, the sun would be a big 'beach ball' about 26 feet (8 meters) in diameter and would be more than half a mile away!

"To take things out another step, we have to adjust the scale. Let's take our beach ball sun and shrink IT down to the size of the orange. Then Jupiter, the largest planet in our solar system, is about the size of a cherry pit, orbiting at a distance of some 260 feet from the orange, while Pluto, the most distant planet, is a mere grain of sand orbiting 1050 feet (three and one-half football fields) away! (If you have a big open area near your school, it would be a fun class project to build a scale model solar system!) On this same scale, the *next nearest star*, which in reality is 4.3 light years (or about 25 trillion miles) away, is another orange-sized object at a distance of 1460 miles! (That's about the distance from Baltimore, MD, to Dallas, TX!). — Bill Blair, fuse.pha.jhu.edu/~Kstanek/

"How Do I Know I Understand This Physics Concept?"

I am probably not the only person who, after days of cramming, managed to get a high grade on a physics exam, and then asked him- or herself, “But do I really understand this stuff? What does it mean to *understand* it?” I was all too aware of the tricks I had used to memorize formulas and proofs, and I knew all too well that within days, weeks, months, certainly years, most of what I had learned would be forgotten (I was not planning to be a physicst). So the standard answer that professors gave to my questions — “Your understanding is measured by your homework and exam grades” I always thought was — if not rubbish, then naive in the extreme.

As the years went by, I felt that “understanding” was too loaded and imprecise a term. If an undergraduate and a physics graduate student both took the same exam and both got As, could you really say that their understanding of the material was “the same”? Of course not. Furthermore a physicist could say to almost any student, regardless of his performance on exams, “Yes, but you don't *really* understand this,” and he would be right, since there are different degrees of understanding.

After much thought, I decided that I should really back way up and ask, “What does undergraduate physics *cover*?” It seemed that that was one way to begin at the beginning. I decided that my answer had to include:

(mechano)-statics
(mechano)-dynamics

electrostatics
electrodynamics

magnetostatics
magnetodynamics

electromagnetism

Special Relativity

That was at least *a* Big Picture. Already the subject seemed a little easier, a little less intimidating. I felt it was at least possible to get my arms, or my mind, around the entire subject.

Then I asked, “What are all technical subjects really about, abstracting away from all the details and difficulties?” Answer: in the last analysis they are all about objects. “And what characterizes objects?” Properties. (That’s really the end of the story, apart from the details.)

And now it was easy to see that a great deal of the content of physics courses was simply the long, wordy, and mathematical, presentation of objects and of ways to find the properties of these objects. What are the properties of matter? Well, among others, mass, density, specific gravity, conductivity, temperature, atomic and molecular structure, velocity, gravitation... What are the properties of an electric field? Its description as a vector function (force per unit charge at each point), geometric description, momentum, energy... What are the basic properties of the object known as the vector cross product? Etc.

By this time I was working on my book, *How to Create Zero-Search-Time Computer Documentation*[1], which presents a method for creating computer documentation that enables users to

1. Available online on the web site www.zsthelp.com

find the instructions they want in less than 25 seconds. Here, the "objects" are the "things" that the user can perform operations on. Thus, for example, for the object "file", the operations might include, generate a file, modify a file, move a file, delete a file, find the properties of a file, upload a file to a web site, etc. I called the documentation for a piece of software an "Environment". William Curtis then generalized the ideas in my book to all technical subjects, and wrote *How to Improve Your Math Grades*[1]. Rapid look-up-ability is the major goal in both books. Curtis's book shows that a complete Environment for undergraduate physics — say, having the content of *Feynman's Lectures on Physics*, would make it possible for any possessor of such an Environment to answer questions like the following:

Quick: what is the definition and value of the constant ε_0?
Quick: what are the basic properties of a magnetic field?
Quick: what are Maxwell's equations in all standard forms (words, differential, integral)?
Quick: what are some standard ways of solving Maxwell's equations?
Quick: what is the rule governing partial derivatives in vector cross products?
Quick: what are all the different types of energy dealt with in undergraduate physics?
etc.

"Quick" here means "in less than 25 seconds". You can make these responses in that short time using an Environment. (By the way, someone is going to make a fortune producing Environments not only for physics but for all technical subjects.)

One of the assumptions underlying the Environment concept is that it is *crazy* to present technical subjects *only* in linear form (as Fenyman's *Lectures* do). The best metaphor I can think of for doing this is: giving a course on the roads and streets of the San Francisco Bay Area in which only the professor has a map. So the job of the students is to come up with their own map, based on purely verbal descriptions of roads and streets and their relationships. Hard work if the students know nothing about the Bay Area! How much easier it would be for them if they had the map when they started (in which case they might not need the lectures, or at least most of them). An Environment is just such a map.

Naturally, Environments lend themselves readily to tests against standard classroom- and textbook presentations of technical subjects. In response to skeptics, I quote my mentor in computer science (John Allen, author of *The Anatomy of LISP*): "You do it your way and I'll do it mine, and then I'll race you around the block."

How Much Will Science Disturb the Universe?

Imagine that mad scientists in some civilization decide that the first order of business is to compute π to as many decimal places as possible, because, after all, π is used in many scientific calculations. Let us assume that there are 10^{87} atoms in the universe, and that the scientists find a way to store one decimal digit in each atom, thus allowing them to expand π to 10^{87} decimal places. (How they keep track of which atom's decimal digit belongs in which place in the decimal expansion we ignore for the moment.)

1. Available online on the web site www.occampress.com

Clearly the universe *after* this colossal calculation, is different from the universe *before* it. In fact, it is hard to imagine what the scientists can hope to accomplish now that they can only make experiments that are guaranteed not to change the decimal digit that each atom stores.

An extreme case, certainly. But science *does* change the universe to some degree, and I am not referring here to measurements at the quantum level. Let us begin by attempting to arrive at some estimate of how much science has changed the planet Earth as of the year 2005. The total number of printed textbooks and journal papers is a beginning, but then we must add the total amount of computer hardware used to store scientific data, and the total amount of materials of all kinds used in scientific experiments throughout the world — in laboratories, linear accelerators, astronomical observatories — and the total amount of materials needed to produce these materials, including mining and transporation and processing equipment (steel mills, aluminum plants, etc.). I think the reader will agree that I have just scratched the surface.

Suppose the real truth about the Theory of Everything is that there isn't enough room in the universe to hold it. Suppose one day physicists will be confronted with the following dilemma: if we carry out this experiment, there won't be enough room to store the results, or the theory predicting the results. On the other hand, if we don't make the experiment, then there will be enough room for the theory predicting the results but no way of determining if the results conform to the theory!

A Small Problem if Time and Space Turn Out to be Composed of Particles

In the theory known as "loop quantum gravity", time and space are quantized, that is, conceived as being composed of particles. At the very least this is a shocking idea for calculus students, since two assumptions that have existed since the discovery of the subject in the late 1600s are that time and space are continuous.

To the thoughtful, though naive, person, a problem immediately presents itself, namely, if time and space are composed of particles, what lies between the particles? In the case of time, it can't be time, and in the case of space, it can't be space. To the person with a little mathematical sophistication, the thought might come to mind: suppose the notion of particles having to exist "in" something is a naive one arising from our concepts of molecules, atoms, electrons and other sub-atomic particles. Can particles be defined by purely mathematical properties that require no something in which the particles exist?

Does Intelligence Matter on the Cosmic Scale?

We contemplate the vast universe, with its billions of galaxies. We contemplate the remarkable discoveries of science in the modern era. And some of us cannot help wondering if intelligence has any importance on the cosmic scale. Certainly we now know things we didn't know in the past — we know about the *existence* of things that no one even thought of in the past. Certainly we are witnesses every day to what scientific knowledge is capable of — for example, enabling us to design machines (e.g., those that release carbon into the atmosphere) that within a few decades may change the planet irrevocably. So we ask: how is the universe different if beings with intelligence at least as advanced as ours, exist for billions of years, as opposed to the universe in which such beings never exist?

"The Trouble With Physics"

If I were a high school student who was doing well in my mathematics and science courses and I read Lee Smolin's book *The Trouble With Physics*[1], I would definitely decide *against* a career as a physicist. The reason is, in a word, that I would not want to spend my life in a vast, fad-driven bureacracy whose ultimate purpose is the generation of Nobel Prize winners. For that is what the profession seems like as Smolin, a respected physicist, describes it. He makes clear, and I think correctly, that nowadays, you either become a Team Player or else you have little chance of getting tenure, or even of getting a job. When we contrast the situation of the young physicist today with that of some of the greatest physicists, and other scientists, of the past, it is almost as though we are not comparing the same professions. The two greatest physicists and the greatest biologist — Newton, Einstein and Darwin — did not attain greatness by falling in line with the dominant thinking of their times (string theory at present in physics) and grinding out papers at a rate set by others. Newton made his first great discoveries while he was living at his mother's farm after Cambridge was closed because of the plague (1665-67); Darwin worked at his home in the English countryside (1838-1882) (he was never on a university faculty); Einstein was a clerk in the Swiss Patent Office while he developed the ideas that resulted in his four great papers in 1905. Smolin makes clear that there are a handful of original thinkers who can somehow survive even in the high-pressure, follow-the-leader physics culture of today, but I for one would have no interest in spending my life in a battle with others about what to think about. I suspect that there are young men and women who feel the same.

A Truly Bad Popularization: Penrose's "The Road to Reality"

There is certainly no lack of useful popularizations of physics. But all of those that I am familiar with, treat the subject at a superficial level — and wisely so, since that is the level that is understandable by the largest readership.

Unfortunately, to write a popularization of greater depth is more difficult than most authors who make the attempt, realize. It is emphatically not a matter of simply covering more material, using prose and illustrations to attempt to soften the difficulty for the non-expert. A case in point is Roger Penrose's 1099-page *The Road to Reality: A Complete Guide to the Laws of the Universe*[2].

At first sight, the book seems to be that miracle of miracles, an understandable presentation of the depths of modern physics. There are lots of illustrations, each inviting in its simplicity. The prose seems, at times, to be directed at an audience of educated laymen. Penrose even states, in the Preface, that the book can be read by persons with little knowledge of mathematics. But my opinion, after weeks of struggling with the book, is that that is utter nonsense, and is one of the reasons why I think the subtitle in full should read, *The Road to Reality...by Someone Who Is Not in Touch With Reality.*

Let me begin with the illustrations. On the one hand, most of them are simple and easy to

---

1. Smolin, Lee, *The Trouble With Physics*, Houghton Mifflin Company, N.Y., 2007.
2. Alfred A. Knopf, N.Y., 2005.

grasp visually. On the other hand, it is almost impossible, in all too many cases, to understand what they are conveying. In some cases, they are frustratingly incomplete — for example, the two depictions of the curves representing sub-series that approximate given series on pp. 78 and 80. We are shown the curves plotted on normal cartesian coordinates, but with *no labels*! We cannot tell, from the graphs, the *x*-*y* coordinates of points where the curves intersect or have maxima or minima, much less the equation of each curve.

Or consider his presentation of the Möbius strip as an example of a fiber bundle (pp. 329-331). It is fair to assume that every reader of the book knows what a Möbius strip is and how to make one (take a band of paper, cut it all the way through, twist the end 180° and re-attach it. A basic property of the strip is that if you move a point along the center, you eventually move it along both sides and return to the starting point. Now look at Fig. 15.5 (p. 331) and the accompanying text and try to figure out how the two cut circles result in such an object.

The extreme of the author's bad habit of not labeling illustrations is unquestionably his diagrammatic tensor notation, pp. 241-242. I think it might be a sign of *mental illness* to publish such hieroglyphics without the most detailed, formal, explanation of what each of the symbols stands for.

The Great Man who wrote the book has spent his life in a culture in which an extraordinary amount of knowledge is in the air. The concepts presented in the book are part of everyday discourse in that culture. And so he thinks nothing of going on for pages about the numerous subatomic particles without providing a table (like the periodic table of the elements) or tree graph to show how they are related. What is the reader to make of these discussions? They are insider talk presented to outsiders in the pathetically naive belief that, because they are in prose, they will be more readily understandable. This explains, I am sure, why, in his treatment of Maxwell's equations, the equations as they are presented in virtually every undergraduate physics book — do not appear! Instead, we are given versions in a strange symbolism that can be understood by insiders only.

It is abundantly clear that the book was never tested on randomly-selected members of its intended audience. In fact, it is abundantly clear that the author never bothered to specify, in detail, before he began writing, the minimum knowledge he was assuming among readers. And that is why we get introductory chapters on the nature of number and on basic calculus, the author having failed to recognize that readers who actually need these chapters, will never be able to understand the sophisticated mathematics that follows, whereas readers who are able to understand, up to a point, the sophisticated mathematics, have no need for the introductory chapters.

Further evidence of his fundamental not-getting-it about what he is doing, is his habit, in the midst of the most abstruse, nearly-incomprehensible, chapter, of pausing to carefully explain that "$n!$" is pronounced "*n* factorial" and that it stands for $n(n-1)(n-2)...1$ — an item of mathematical knowledge we learn in high school.

We can be very sure that, instead of testing the book on readers, the author merely circulated parts of the manuscript among colleagues and other persons in the physics community. These readers knew most of the subject matter like the back of their hand, and so, naturally, the text seemed crystal clear. And, of course, the know-nothing staff in the publisher's office assumed that, if the Great Man's colleagues say the book is clear, why, then it must be clear (even if it is bewildering to the members of the staff).

One of the worst examples of mathematical popularization I have ever come across is Penrose's chapter on fiber bundles ("Fiber bundles and gauge connections", chapter 15). Does the author, or the publisher, or any person who believes this chapter is "instructive", have the slightest idea of its level of comprehension across the intended audience of readers? This is little more than mathematical dazzlement, the reader telling him- or herself, on the basis of the occasional understandable sentence and the illustrations, "Well, certainly this is clear. Now if only I had time to figure out what it means..."

Finally, a word about the index. It is clear that, like all technical authors, Penrose considers an index an afterthought, when in fact it is one of the most important parts of a technical book. In this one –a thousand-page survey of modern physics, the following terms, among many others are *not* in the index: "relativity", "special relativity"[1], "general relativity", "tensor calculus" (the mathematics that underlies general relativity), "differential", "form", "1-form", "2-form", etc., "two-slit experiment" (the most famous experiment in quantum mechanics).

After weeks of struggling with this disgrace of a book, I can only say, "*If this book represents how a great physicist thinks, then thank God I am not a great physicist!*"

How Popularizations Should Be Written

The task must begin with a clear definition of the audience, one that is based on minimum skills and knowledge that members of the audience are expected to have. Not, for example, "high school mathematics", or "at least two years of college engineering or physics", but a list of specifice courses, for example, "high school algebra through solutions of quadratic equations", "at least two semesters of elementary calculus", etc.

Next must come answers to the crucial questions: What are the goals of the book? and What tests will be accepted as indicating that the book is achieving those goals?

Before beginning, the author must purge himself of the embarrassingly naive belief that prose makes technical topics easier to understand by readers without strong technical backgrounds. Rather, he or she will take as his working goal, the making of each concept as *rapidly understandable* as possible. In many cases this will mean using tables — for example, in any discussion of particles, the author will have the periodic table of the elements always in mind. Of crucial importance is an indented list or tree graph showing the types of each entity that a chapter is about, e.g., types of field, space, bundle, manifold, etc. and how these types are related to each other — which is a sub-type of which, etc. Nothing makes a chapter more rapidly understandable than having this kind of Big Picture to look at.

Mathematical statements will always be presented in the same format, for example, If ... and ... Then ..., with the logical terms in, say, boldface, and the logical phrases always indented in the same way. (The author must carefully read the chapter, "Proofs", in William Curtis's book, *How to Improve Your Math Grades* on occampress.com.)

When a new concept is introduced that has several parameters, e.g., fiber bundles in the case of Penrose's above-mentioned book, then the list of parameters should be repeated in the *same*

1. Incredibly, unbelievably, there is virtually no mention of special relativity in the entire book — a theory, presented by Einstein in 1905 that was, along with Planck's theory of quanta , the beginning of 20th century physics, and, in particular, the predecessor of Einstein's general theory of relativity, about which there is likewise virtually no mention.

table form for *each* example, because this is a great aid in speed of understanding. Thus in the case of fiber bundles, this form should be:

Name of bundle:...;
M (base space):...
V (fiber):...
Canonical projection map:...
Cross-sections:...
...

Each parameter can then be elaborated upon below the table.

My feeling at present is that unless the audience is primarily people with a mathematics background, all or most equations and other formal mathematical statements should be relegated to one or more appendices, along with their proofs when proofs are given.

As each chapter is completed, the chapter must be tested on randomly-selected members of the intended audience.

As the book proceeds, a complete, thoroughly cross-referenced index, *including symbols*, must be generated.

Standard notation — the most commonly-used notation in existing, popular textbooks — must be used throughout.

There shomuld be illustrations to make difficult concepts easier to understand. However, expcept in the simplest cases, each illustration should have a sequence of circled numbers to indicate *the order in which the illustration is to be looked at*. News magazines have done this for years. For some reason, the value of the idea has so far escaped popularizers. Illustrations, like text, must be tested for comprehension.

With these practices, the author has a fighting chance of producing something beyond a mere exercise in vanity.

\

Computer Science

Why Is Computer Software Still So Difficult to Use?

It is nothing short of a disgrace that in 2015 computer users are still at the mercy of software designers who, when judged by the proper criteria, are all but incompetent. (*Note*: I exclude Apple software from this criticism because, although I have never used it, I have heard it praised sufficiently by users to regard it as an exception to standard industry practice. The probable reason for Apple's success at software design is given in the last paragraph of this section.)

Let me take a product that represents a collection of the all-too-frequent bad practices in software design. It is FrameMaker 11.0, a word-processor (costing more than $1,000) that is widely used by software documenters. One of its predecessors, FrameMaker 7.0. which I used for several years, was quite acceptable, but in my opinion FrameMaker 11.0 is a disaster. If you were unfortunate enough, as I was, to buy the product soon after it was released, what you got was a product with far too many bugs; in addition, the marketing geniuses decided that one way to enhance the impression that this was a new product was to change the screens so that a person long familiar with Frame in effect had to relearn how to perform basic operations he or she had performed for years.

Furthermore there was absolutely no help provided in troubleshooting. You had no choice but to hire a consultant. There was no central company-sponsored web site where one could search for recommended fixes for problems. If by chance you were clever enough to somehow get the email address of Support, you could sometimes get a little guidance from personnel who, let it be said, tried to be helpful, but usually knew far too little about the problems users were running into.

In addition, whereas in Frame 7.0 it was possible to access *all* files generated by earlier versions of Frame, now this was not possible: some of the old files could not be accessed at all, others only with long complicated work-arounds involving other manufacturers' software.

I need hardly mention that the designers completely ignored the well-known practice of, e.g., Windows, of downloading, to each user's computer, revised versions of the software that contained fixes to bugs.

Finally, the documentation showed the one sign that is a sure give-away that it was created by people who had never thought deeply about how documentation is used, namely, the belief that more is always better. In fact, there was no quick way even to get a *list* of all the available documentation and training programs. There was a well-written manual, that is true, but it did not provide an overview of all the available documentation and training that was available, much less a recommended way to proceed through — to use — this wealth of material.

So let me now lay out a design protocol that would have avoided all these problems.

• First and foremost, a written-out description of what each user will be assumed to know before starting to use the product. *Note*: If the head of Marketing ever says "This product will be usable by *anyone*", show him to the door.

• Absolutely no human interface changes over the previous version of the product if the purpose is merely to make the new product look "new". The belief that making cosmetic changes to the human interface increases sales is hopelessly naïve. Those who have not used Frame before will not know the difference. Those who have will be *much* more impressed by ads that say, in effect, "the screens for the old functionality are exactly the same in the new version. But in addition, there is *new* functionality!..."

• Outside-In Design — the human interface should be designed *before* the supporting programs, not after! This, of course, should be a top-down process:

(1) write the top level human interface based on the *tasks* that users can perform at this level;
write the top level only of supporting programs for this level,
test on prospective users; then

(2) write the second level human interface based on the *tasks* that users can perform at this level;
write the top level only of supporting programs for this level,
test the top *two* level human interfaces on prospective users, then ... etc.

Complete the writing of all the supporting programs only after all the human interfaces have been written and tested.

This process of course demands that the documenters be brought in *on the very first day*, because it is entirely possible, in fact, likely, that the user will want to or need to refer to the Help facility at the top level. As it happens, a method has been worked out that offers documentation that enables users to find the instructions they want in less than 25 seconds at least 80% of the time. It is called “Zero-Search-Time” (zst) documentation. For details, see zsthelp.com. (I am the author of the method.)

• Automatic updates (in the manner of, say, Windows) of each user's copy of the software throughout the life of product, thus enabling bugs to be fixed after product release.

.

• A web site containing an indexed list of all known problems with the software, and recommended fixes, the web site address being made known to all users, including via the zst index in the Help facility and via notices in all other documentation.

And last, but not least,

• Profit-sharing for all members of the project team.

Let me conclude by trying to explain why Apple software is so much better than the typical industry product. The reason, in my opinion, is that Steve Jobs kept a tight rein on the programmers at all times. He seemed to understand from the beginning that if you assume that programmers know how to create a product, you will get, not something that is usable by the intended class of users, but instead a collection of demonstrations of how brilliant the programmers are. The attitude of modern-day programmers is still what it was in the early days of computers: “If you need help in learning how to use my program, then you have no business using it.” God knows how much in time and effort and frustration and misery this attitude has cost users over the years. But Jobs knew it was not to be tolerated.

A Frequency-of-Execution Approach to Programming

We can regard a compiled computer program as one that assumes that no statements will change between executions. We can regard an interpreter as a program that assumes that any statement may change between executions. Why not design programs on the assumption that frequency-of-execution cannot be predicted at all? Then, in actual use, the most frequently-executed statements will be represented as tables of input-output pairs (or tuples, in the general case). Periodically, based on the statistics of actual program execution, the program will change what gets tabularized in this fashion. The programmer can set an upper bound to the amount of space to be reserved for these tables and to how often data on frequency-of-execution is collected and how often the tables are revised. Since table-look-up is in general much faster than computation, a program with such a table should be much faster than the same program without the table.

Can we consider a program as an information source — one that, during execution, transmits assembly language instructions to the cpu (central processing unit), the latter acting as the receiver? How would interpreter and compiler design be different if this were the model of computation the design was based on?

"Calculation Is Successive Classification."

Exercise: Discuss the statement, "Calculation is successive classification." Show how various programs can be viewed as carrying out a classification process. For example, when adding two numbers, we successively classify the sum as, first, being one that has the digit we get in the 1s place (the sum mod 10), then the sum as one that has the digits we get in the 1s and 10s place (the sum mod 100), etc.

Discuss: "I can't think. I can only classify."

"Backwards" Grammars

Discuss the kind of formal grammars that result if, for a given grammar, we reverse all the arrows in the productions defining the grammar. How are the two sets related, i.e., what types of functions exist between the set of formal grammars and the set of "backwards" formal grammars?

Grammars as Information Sources

Consider a grammar as an information source. By reasoning, or by many trials, develop statistics on the probability of occurrence of each symbol in a terminal string. Can these statistics then be used to determine the probability that a given string is in fact in the language generated by the grammar? Can this idea be applied to probabilistic proofs of theorems, so that we can make statements such as, "The probability that the statement (string) x in theory (language) y is a theorem is z"?

Syntax vs. Semantics

Are we really clear on the difference, in a programming language context, between syntax and semantics? Should we be looking at the matter topologically, and speak in terms of "distance", so that things become "more syntactic" relative to a given starting point as the distance from that starting point increases? Or is it rather that semantics just refers, ultimately, to certain kinds of partitions? After all, I need syntax to describe the pure function I am talking about. I need rules for expressing what I mean. But I have trouble believing in a concept as an airy something, a pure form in a Platonic heaven somewhere. Or maybe semantics just means, "any of these syntactic representations, and preferably the simplest", and syntax means, "just this representation". We should be aware of being led astray by familiar examples, e.g., those of ordinary arithmetic, with which we are so familiar that it does seem that there is a somewhere that contains the pure idea. But suppose we are dealing with extremely complicated functions, about which we are not at all sure how to get a "simple" syntactic representation. Then, wouldn't we start aiming for the ability to recognize whether a given supposed representation of the function was in fact a representation? I.e., wouldn't we be groping for quick ways to decide if this piece of syntax could be mapped into one we knew represented the function? Wouldn't that ability begin to feel like an "intuitive understanding"? Or, to put it another way, since syntax, as well as semantics, are ultimately expressed in symbols, what distinguishes the two, on a purely grammatical basis?

Given a binary tree with each pair of branches at each node being marked "0" and "1", we can specify any binary integer by giving a path down through the tree, beginning at the root. Each path is unique, and each specifies a unique integer if we agree that the binary digits in each path are in the order of most significant to least significant digit, and if we agree that there are to be no leading 0s.

One problem with proving the correctness of computer programs is that there is an infinity of programs to compute each function (we can just add null statements to any program that computes the function).

So, unlike the case of our tree to generate binary integers, there are many "paths" (programs) leading to each function.

If we could somehow overcome this, so that, the "path" would itself define the function, as the path itself defines the integer expressed in binary notation in our tree, then the problem of proving correctness would become trivial.

A Question About Information Theory

Does Shannon's Theorem answer the coding question in its full generality, i.e., the question: given a known probability of error of successfully transmitting a signal, and furthermore, given a description of the error-causing background, are there better and worse codes to use? E.g., how about a linearly rising voltage for 0, a linearly falling voltage for 1. In other words, can the symbols we choose to represent the symbols *themselves* affect reliability of transmission?

Finite-State Machines Viewed Graph-Theoretically

A finite-state machine is a Turing machine with only a finite amount of memory, i.e., only a finite number of tape cells. Call the triple <state of finite control ("program"), tape contents, tape head location> a *configuration.* Then, since there are only a finite number of states of the finite control, and only a finite number of tape cells, hence only a finite number of possible tape head locations, there are only a finite number of possible configurations. Therefore, there exists a graph in which each node is a configuration. There is an arrow from each node to the next node, as dictated by the finite control. Do we gain anything from this view of a finite-state machine?

Against Viewing Information as Something That Is "Sent"

If I were asked to predict one of the next major advances in communication, at least in communication between computers, and between parts of computers, e.g., between the cpu and the graphic display, then I would say the advance will be in the slow erosion of the idea that information is something that is "sent" from one place to another, just as I would predict that one of the next major advances in computer science will be the slow erosion of the importance of "sequence", because this will be replaced by a kind of "simultaneity". For example, instead of looking at computations as sequences, we will look at all possible computations by a given program at once, and then merely talk about points in this computation "field". Now in communication, I also want to get rid of "before" and "after" and instead look at simply an entire "field" consisting of the bits at the source, the bits in the transmission line, and the bits at the destination *as one single thing that changes over time*. If someone wants to view the fact that at time t_k a string of bits was at location $<x_1, x_2, ..., x_{k>}$ in a transmission line, and at time t_{k+1} the string was at locations $<x'_1\ x'_2, x'_3, ..., x'_{k+1}>$— if someone wants to call this the "movement" of the string because it is being "sent" from the source to the receiver, that is fine, but I prefer to look at it as merely how the entire field of bits consisting of those in the source, those in the transmission line, and those in the receiver, happens to have changed.

To take a more familiar example, at least one that is familiar to programmers: it is rare that every bit in a graphics display changes with each click of the cpu clock. Typically only a few bits change, so that, if someone wants to speak in terms of the image being sent each time, well, that's fine, but we know that what is really happening is simply that a given image is being modified slightly. We are sending the changes only.

Computing How Long It Will Take to Compute How Long It Will Take To...

We have some computation *x* to perform. But before we actually begin the computation, we want to find out how long it will take. So we have a computation *y* to perform. But before we actually begin that computation, we want to find out ... *Project*: investigate such sequences of computations with a view of finding out a rule for when to end the sequence.

"Graphing" Partial Recursive Functions, or, "An Algorithm Does Not Exist...": Why?

In analysis we study functions of complex numbers that for certain arguments yield infinite values. These arguments are called "singularities". We can graph at least portions of such functions, and show where the singularities are, and how the function behaves in the vicinity of each.

In computation theory, we study functions — the partial recursive functions — that for certain arguments, are undefined. Given such an argument, a program that computes the function simply keeps computing forever. Can we create a graph that will enable us to see how undefined values arise — that will show us what inputs are "almost" undefined, i.e., that will show us the behavior of the function in the vicinity of each undefined value?

Without loss of generality, we can limit our discussion to functions of one argument (input). Let *T* be a program (Turing machine). With each input *x* (finite binary string) to *T*, we can associate the number *n* of instruction executions that are required to arrive at a value if the computation halts. Thus, for *T* we have a set *S* of ordered pairs $<x, n>$. Now sort the elements of *S* on *n*. The result will be the subset (possibly empty) of all $<x, n>$ such that $n = 1$, and the subset (possibly empty) of all $<x, n>$ such that n = 2, etc.

Such a sorting might help us answer questions like, "What are the characteristics of inputs that are 'near' to inputs that are undefined?"

.Is there a topology on the partial recursive functions that allows us to make informal statements of the sort, "Now here, in this region, the total functions begin to become partial."? The answer would seem to be "no", because either a computation halts or it doesn't. On the other hand, if we mean by "to become", having an increasing number of arguments for which the computations do not halt, then perhaps. But once we are in the region in which all the functions have an infinity of such arguments, then it would seem we cannot speak of this number "increasing".

Continuous Design

As computer memory and computer speed keep increasing, we may be approaching the point where we can employ what might be called the "continuous" design of physical objects. For example, suppose we want to design an improved high-performance sailplane. We represent in the computer a current high-performance sailplane, then allow a program to continuously distort the plane by continuously changing all parameters. Simulated wind-tunnel tests can be performed during the process as desired. When the optimum shape is reached, the program terminates. Are there better (more efficient) and worse ways to perform such a continuous process?

A Non-Self-Referential Proof of the Unsolvability of the Halting Problem

Turing's Self-Referential Proof

Let *L* be a programming language in which a program for each computable function can be written. Let *S* denote a program written in *L* and let *I* denote an input to the program *S* (without loss of generality, we can assume that each program has exactly one input). Then the Halting Problem asks if there exists an algorithm (program) *P*(*S*, *I*) that will return "S(I) halts" if *S* eventually halts in computing *I*, and "S(I) does not halt" if it doesn't.

Turing's proof that no such algorithm exists is, in brief, to assume that, on the contrary, such an algorithm, *P,* exists, then to show that this implies that there exists a program *Q* that halts if *Q* doesn't halt, and doesn't halt if *Q* halts, which is obviously a contradiction.

The program *Q*(*S*) is as follows:

1. Compute *P*(*S*, *S*).

2. If *P*(*S*, *S*) returns "*S*(*S*) halts" then make *S*(*S*) go into an infinite cycle (i.e., never halt).
   If *P*(*S*, *S*) returns "*S*(*S*) does not halt" then make *S*(*S*) halt.

Now let *S* = *Q*. Then the computation for *Q*(*Q*) is as follows:

1. Compute *P*(*Q*, *Q*).

2. If *P*(*Q*, *Q*) returns "*Q*(*Q*) halts" then make *Q*(*Q*) go into an infinite cycle (i.e., never halt).
   If *P*(*Q*, *Q*) returns "*Q*(*Q*) does not halt" then make *Q*(*Q*) halt.

The claim is that

if *Q*(*Q*) halts, then it doesn't halt;
if *Q*(*Q*) does not halt, then it does.

This contradiction, it is claimed, proves there is no such algorithm P.

What bothers me about the proof is that the program Q is self-modifying, and so we cannot speak of *the* program Q. In step 1 in the computation of $Q(Q)$ (see above), there is no contradiction. If $Q(Q)$ halts, then the program Q is changed in a certain way. If $Q(Q)$ does not halt, the program Q is changed in another way. The contradiction arises when we call the original Q and the changed Q by the same name.

What we really have is an infinite sequence of programs Q: Q_0 when we start, then Q_1 after the modification has been made. Q_1 would be the Q on which $Q(Q)$ operates if we were to run Q again. And so forth. All Q_i with even subscript would behave the same; and similarly for all Q_i with odd subscript. It is only in going from Q_i to Q_{i+1}, where i is even, that we get the claimed contradiction.

Is the real contradiction the fact that the "same" program changes its behavior with repeated executions?

Another infinite sequence is the following. From our description above of the computation of $Q(Q)$ we can write $Q(Q) = f(Q(Q))$. But that implies that $Q(Q) = f(f(Q(Q)))$, which in turn implies that $Q(Q) = f(f(f(Q(Q))))$, etc.

Non-Self-Referential Proof of Unsolvability of Halting Problem

1. *Definition*: a semi-decidable set is one having an associated program such that, if the program is given an element *s* that is in the set, then the program will halt with the equivalent of the message, "Yes, in set." However, if *s* is not in the set, then the program may never halt. Furthermore, there are no associated programs that are guaranteed to halt whether or not *s* is in the set.

2. Semi-decidable sets exist. For example, Type 0 languages in formal language theory are semi-decidable sets, since for each such language, a program (called a "recognizer") exists that will halt if given a string *s* in the language, but may not halt if *s* is not in the language. Furthermore, there are no associated programs (recognizers) that are guaranteed to halt whether or not *s* is in the set.

Semi-decidable sets have always existed, and always will exist. This is true even if no one had thought of the Halting Problem, much less found a proof for its unsolvability, or even if no one had defined these sets. Turing's 1936 proof (above) of the unsolvability of the Halting Problem did not somehow "call into being" semi-decidable sets. A proof does not call entities into being, it merely establishes one or more facts about them. Similarly, a definition does not "call into being" an entity, it merely calls attention to it, and gives it a name.

3. Since semi-decidable sets exist, by definition there does not exist an algorithm to determine membership in such a set. But if the Halting Problem were solvable — that is, if a program Q existed that, given any program and any input to that program, would return "Yes" if the program halted on that input, and "No" if it did not — then there *would* be an algorithm to determine set membership for semi-decidable sets. For we could simply give Q the program P associated with the semi-decidable set S, and any input s to P, and if Q returned "Yes" then we would know that s was in S[1], and if it returned "No", then we would know that s was not in S. But no algorithm can exist to determine membership in a semi-decidable set, and so the Halting Problem is unsolvable. □

1. It is possible that, in some cases, P might return "No" if a string was not in the set. But by definition of semi-decidable set, it could not do this in all cases.

Remark

The proof would be circular if semi-decidable sets had somehow been "called into being" by Turing's 1936 proof. But the proof brought no such set into being. (See step 2 of our proof above.)

If the Halting Problem Were Solvable...

If the Halting Problem were solvable, the $3x + 1$ Problem would then be decidable. This Problem asks if, for all inputs, the computation of the $3x + 1$ function terminates in 1. If it does not, then it is known that the computation never terminates for that input. The existence of the program *U* would enable us to know, for any input, whether the computation terminates in 1 (the computation halts) or if it does not (the computation does not halt). At present, if the computation has not terminated in 1 by the time our compuation resources of time or storage space run out, we do not know if the reason is that the computation requires more resources than we currently have at our disposal, or if in fact the computation never terminates.

If the Halting Problem were solvable, Fermat's Last Theorem (FLT) would also be decidable. Consider the program *V*:

for all $k \geq 3$ and for all $x \geq 1$ and for all $y \geq 1$ and for all $z \geq 1$ do
begin
if $x^k + y^k - z^k = 0$ then halt, else continue
end

The program *V* would be given to *U* , the program that solves the Halting Problem. (The input would be the null input) If *U* halted with "The program halts" then that would mean that a counterexample to FLT exists and FLT is false; if *U* halted with "The program does not halt" then that would mean that FLT is true.

Appendix A — Obscurities and Omissions in Einstein's First Paper on Special Relativity

Before I begin, I want to make crystal clear that I am *not* disputing the correctness of Einstein's results! I am only pointing out obscurities and omissions in Einstein's paper, "On the Electrodynamics of Moving Bodies", which can be found in pp. 37-65, in Einstein, A., Lorentz, H. A., Minkowski, H., and Weyl, H., *The Principle of Relativity* (Dover Publications, Inc., Mineola, N.Y., 1952).

Section 3

Section 3 is found early in Part I of the paper.

Einstein defines two frames of reference:

K, with coordinates x, y, z, t is the fixed frame of reference;

k with coordinates ξ, η, ζ, τ. is the moving frame of reference.

The two frames are initially superimposed on each other, i.e., the axes are parallel and the origins coincide. Then k begins moving to the right (in the direction of increasing x in K) at the velocity v.

Einstein says: "If we place $x' = x - vt$ it is clear that a point at rest in the system k must have a system of values x', y, z independent of time."

It is not at all clear what this sentence means. In other sources, e.g., Feynman[1], Vol. 1, p. 15-2,. $x' = x - vt$ usually represents the following:

Pick a point x in K, the fixed frame of reference. Then the point x' in k, the moving frame of reference, that corresponds to x (is "on top of" x) at time t in k is given by $x' = x - vt$. But this point x' changes as t changes, because k is moving. There is an infinity of x' corresponding to a fixed x. In other words, Feynman and other sources hold that: for each x and for each t there exists an x' such that $x' = x - vt$.

(We can't help noting that, since x' is a point in k, the moving frame of reference, it should be written ξ'.)

Einstein goes on to say, "From the origin of system k [the moving frame of reference] let a ray be emitted at the time $\tau_{.0}$ along the X-axis to x', and at the time $\tau_{.1}$ be reflected thence to the origin of the coordinates, arriving there at the time $\tau_{.2}$; we then must have $1/2(\tau_{.0} + \tau_{.2}) = \tau_{.1}$ or, by inserting the arguments of the function τ and applying the principal of the constancy of the velocity of light in the stationary system: —

(1)

$$(1/2)\left[\tau(0, 0, 0, t) + \tau\left(0, 0, 0, \frac{x'}{c-v} + \frac{x'}{c+v}\right)\right] = \tau\left(x', 0, 0, t + \frac{x'}{c-v}\right)$$

1. Feynman, Richard P., Leighton, Robert B., Sands, Matthew, *Lectures on Physics*, Addison-Wesley :Publishing Company, Reading, Mass., 1964.

The thoughtful reader is immediately struck by the term $c + v$. What possible meaning can such a term have after Einstein has made clear, earlier in the paper, that c, the speed of light, is the maximum speed in the universe?

Second,

(2)

$$\frac{x'}{c-v} + \frac{x'}{c+v}$$

is the time it takes the light ray to reach the point x' and return to the origin — of which frame? The lower-case italic typeface indicates the fixed frame, K, but the light ray is being sent in the moving frame k. If we assume that the expression is the round-trip time as observed from K, then why is the $c - v$ fraction first, since this fraction represents the time (as observed from K, the fixed frame) to *return* to the origin?

The answer seems to be the following: the expression (2) is the time *relative to k, the moving frame*, for the light ray to travel to x' in k and return. (If v were very close to c, then $c - v$ would be very close to 0, meaning that the light ray would move very slowly (almost at a speed of 0) along the axis ξ. Hence the time to travel to x' would be very large (almost infinite). The round-trip time would be very large no matter how fast the return trip (at a speed of $c + v$).

But since the expression (2) is the time for the light ray's round trip in k, the moving frame, how can it be an argument for the function τ, whose arguments by definition are in K, the fixed frame?

Similarly,

(3)

$$t + \frac{x'}{c-v}$$

in the right-hand side of equation (1) mixes a time t in K, the fixed frame, with the length of time in the moving frame k that it takes the light ray to travel from the origin of k to x' in k. But then how can expression (3) be an argument of τ, which by definition takes arguments in K, the fixed frame?

The algebra that follows equation (1), at least for me, is impenetrable, since it appears to be algebra on both the arguments of the function τ and the value of the function, which has not yet been formally defined.

I should mention that there seems to be an error in the following two statements later in the section. On the one hand, Einstein writes

(4)

$$\xi = a\frac{c^2}{c^2 - v^2}x'$$

and then, a few lines later, he writes

$$\xi = \varphi(v)\beta(x - vt)$$

where, he states,,

$$\beta = \frac{1}{\sqrt{1 - v^2/c^2}}$$

Earlier, he stated that $\varphi(v) = a$, and showed that $x' = x - vt$. But, as the reader can easily verify,

$$\beta = \frac{c}{\sqrt{c^2 - v^2}}$$

and so the term he has in equation (4) is really β^2, not β, and that is an error.

If nothing else, we see, from Section 3, the crucial importance of *always* specifying:

the frame F_1 in which a moving object U (e.g., a light ray) is moving;
the speed r at which U is moving relative to F_1;
the speed s at which U is moving relative to another frame F_2;
the speed v at which F_1 is moving relative to the frame F_2.

Section 6

Another difficulty in the paper occurs in Section 6 at the start of Part II, "Electrodynamical Part". The section begins: "Let the Maxwell-Hertz equations for empty space hold good for the stationary system *K*, so that we have ...". There then follows a list of equations which the reader finds can be represented as

$$\frac{1}{c}\frac{\partial \boldsymbol{E}}{\partial t} = \nabla \times \boldsymbol{M}$$

and

$$\frac{1}{c}\frac{\partial \boldsymbol{M}}{\partial t} = \nabla \times \boldsymbol{E}$$

where **E** is the electric field and **M** is the magnetic field.

But if the reader checks the list of Maxwell's equations in, e.g., Feynman (ibid.), Vol. II, p. 18-2, he does not find these! He finds that the first equation is missing one term and a factor *c*, and that the second equation should be negative on the left-hand side.

Even more perplexing is the mapping from the **E** and **M** coordinates in the stationary system *K* to their corresponding coordinates in the moving system *k*. that then follows. Einstein says only that this mapping is achieved by applying "to the above equations the transformation developed in Section 3". Yet the reader will be hard-pressed to figure out how the mapping follows from anything in Section 3.

The above perplexities in Sections 3 and 6 could be cleared up in an annotated edition of Einstein's paper, thus saving readers considerable time in trying to understand one of the classics of modern physics.

Appendix B — On Basic Tasks in Special and General Relativity

In this Appendix we consider some of the basic tasks that are performed in Special and General Relativity. The goal is to make the definitions of the tasks, *and their implementation*, as precise as possible. The implementation of each task should be described via a sequence of steps. To take what is probably the simplest example, namely, to measure the distance, in a frame of reference, relative to that same frame of reference, to a point x on the x-axis:

1. Place one end of the unit measuring rod at the origin.
2. Move the lower end forward along the axis so that it is at the location where the higher end was, and keep repeating this until you reach the point x. The number of rod lengths is the distance to x.

An explanation must be provided as to how to deal with the possibility that x is a fractional or irrational distance from the origin, or else a separate procedure must be given to overcome the distance to x being fractional or irrational.

We also must raise the question why, at present, we could not legitimately use a light beam, and an atomic clock, to measure the distance. The time it took the light beam to traverse the distance, multiplied by the speed of light, would give us the distance.

If the reader feels that such a procedure for measuring a distance is trivial and doesn't require written-out instructions, let him or her give a procedure for synchronizing two clocks in a frame, or for determining if two events are simultaneous in a given frame of reference, relative to that frame, or for determining the speed relative to a frame F_s of a speed in a frame F_m, where F_m denotes a moving frame, and F_s denotes a fixed frame.

Such function (task) definitions and procedures can do wonders to overcome the false sense of understanding that mere prose explanations induce. In particular, they make clear the necessity of always clearly specifying: (1) in what frame the measurement is being made; (2) who is making the measurement (an observer in the frame, or an observer in another frame?); (3) what the equivalent measurement is in the other frame (Lorentz Transformation); (4) whether a frame can exist inside another frame, and, when this is not the case, what the dividing line is between two separate frames. For more details, see the "Environments Make Relativity Much Easier to Understand" on page 63, and the sub-sections that follow.

Statements Need to Be in Bold-Face Type!

What is needed, in popularizations as well as more formal treatments, is the clear marking of the *statement* that each often long-winded prose exposition is setting forth — **the statement in bold-face type**, preferably *before* the prose exposition. For example, what exactly is being shown, set forth, regarding simultaneity? We need and deserve formal statements, in the standard form of lemma and theorem statements, not just prose arguments that plant uncertainty about simultaneous events in our minds.

Environments Make Relativity Much Easier to Understand

There is a type of presentation that would make all the others (popularizations or not) much easier to understand, *regardless of how much mathematics they contain*. It is the one that William Curtis, in his *How to Improve Your Math Grades* on occampress.com, calls an *Environment*. An Environment is like an encyclopedia, so that any term, including any frequently-occurring algebraic expression, can be looked up alphabetically, or in a section of expressions and symbols at the end. “Whatever *can* be made look-up-able, *should* be made look-up-able”, Curtis says, and I agree with him.

Curtis’s book teaches us that any technical subject can be regarded as a collection of entities on each of which certain operations, or tasks, are commonly performed. Explanations of how each operation is performed, are provided in the Environment. So, in the case of Special Relativity, we have the following.

Some Common Operations in Special Relativity

Students are normally introduced to Special Relativity by one or more examples showing that the value obtained from a physical measurement depends on the location and relative speed of the person making the measurement. But then students are told about the constancy of the speed of light, the shrinking of length, the slowing of time, and the increase of mass as an object approaches the speed of light. It seems to me that scientific rigor demands that a table like the fol-

Table 1:

Location of person making measurement	**Motion of Frame where person is**	**Location of thing to be measured**	**Motion of Frame containing thing to be measured**	**Thing to be measured or done**
Frame F_1	Not known	Frame F_1	Not known	Length (distance between two points)
Frame F_1	Same as for F_2	Frame F_2	Stationary relative to Frame F_1	Length (distance between two points)
Frame F_1	Same as F_2	Frame F_2	Moving at constant velocity relative to Frame F_1	Length (distance between two points)
Frame F_1	Not known	Frame F_1	Not known	Time between two events
Frame F_1	Same as for F_2	Frame F_2	Stationary relative to Frame F_1	Time between two events
Frame F_1	Same as F_2	Frame F_2	Moving at constant velocity relative to Frame F_1	Time between two events
Frame F_1	Not known	Frame F_1	Not known	Mass of an object

Table 1:

Location of person making measurement	**Motion of Frame where person is**	**Location of thing to be measured**	**Motion of Frame containing thing to be measured**	**Thing to be measured or done**
Frame F_1	Moving at a speed v	Frame F_2	Stationary relative to Frame F_1	Mass of an object
Frame F_1	Moving at same speed as F_2	Frame F_2	Moving at constant velocity relative to Frame F_1	Mass of an object
Frame F_1	Not known	Frame F_1	Not known	Synchronize two clocks
Frame F_1	Moving at a speed v	Frame F_2	Stationary relative to Frame F_1	Synchronize two clocks in F_2

lowing be shown to students, with the stern admonition that whenever they (or professors) speak of the making of a measurement, the discussion must be preceded by a row like the ones in the above table.

A more extensive table in a first semester course on Special Relativity would also include the following:

Determine if two clocks in a specified Frame are synchronized relative to a specified Frame;

Determine if two events are simultaneous for a given observer;

Measure the length of a train from the gound outside the train;

Measure the distance between two point on the ground outside a moving train, from within the train;

Compute how much a massive object in a moving Frame is observed to shrink in the direction of its movement, by a person in a stationary Frame;

Compute how much time slows down within a moving Frame, relative to a person in a stationary Frame;;

Compute how much the mass of an object in a moving Frame is observed to increase by a person in a stationary Frame.

Other operations are given in the section, "“What Time Is It?” “Is What?”" on page 39.

In the case of General Relativity, we have:

Some Common Operations in General Relativity

Compute the interval (distance) between two events;

Determine the type of space-time that exists in a given region;
Find the curvacture of a given region of space-time;
Find the path that a particle or object takes in a given region of space-time.

The mathematics that underlies General Relativity is tensor calculus. It, too, can be treated in the Environment format. The list of common operations that are performed on tensors in itself makes the subject more accessible, because, like all such lists, it clearly separates the What from the How — something that the vast majority of presentations, regardless of their mathematical level, usually fail to do.

Some Common Operations on Tensors

Prove that a term is or is not a tensor;
Determine if two tensors are equal;
Add two tensors;
Multiply two tensors;
Determine the type of a tensor;
Contract a tensor;
Lower one or both suffixes of a tensor;
Construct a tensor of higher rank.

To a person with a background in formal language theory, tensor calculus seems just a complicated formal grammar, with most of the common operations easily performed by computer programs. (E.B. told me that there are more than 200 such programs!) One problem with all the presentations of the subject that I have seen is that it is never made clear what tasks in General Relativity the tensor operations implement.

Appendix C — Is There an Alternative to the Lorentz Transformation?

The Lorentz Transformation establishes a relationship between the coordinates x, y, z, t in one frame, and the coordinates x', y', z', t' in another frame. Statements of the Transformation can be found in virtually any Special Relativity textbook and in many popularizations.

Most of the derivations of the Transformation involve spheres of light. It is natural to wonder if the relationship between the two sets of coordinates can be derived from a simpler, more intuitive basis. In this Appendix we attempt to do that. Unfortunately, we find that our results differ from two in the Lorentz Transformation.

Let a frame have the set of Cartesian coordinates, x, y, z, t. We denote the x and y axes, X and Y respectively, and the origin O. We will identify the frame by the name of the origin, O.

Let another frame have the set of Cartesian coordinates, x', y', z', t'. We denote the x' and y' axes, X' and Y' respectively, and the origin O'. We will identify the frame by the name of the origin, O'.

We will only be concerned here with the x and t coordinates because $y = y'$ and $z = z'$. Initially the origin O' is superimposed on the origin O, and the X'-Y' axes are superimposed on the X-Y axes, respectively.

Now suppose at time $t = t' = 0$, frame O' starts moving to the right, along the X axis at a speed v, where $v << c$, the speed of light.

We assert that

(1)

for all t, t', $x = ct$, and $x' = ct'$.

Then at any time $t > 0$, the distance from the origin O to the origin O' along the X axis will be $x_{O'} = vt$.

The distance $x = ct$ will clearly be greater than vt. The difference between $x = ct$ and $x_{O'} = vt$, is x' That is,

(2)

$x' = x - vt$.

Now, since $x = ct$, (2) implies,

$x' = ct - vt$, or

(9)

$x' = t(c - v)$.

This statement shows that for each t there is a unique x'.

Since, from (1), $x = ct$, we have $t = x/c$, so we can write

(10)

$$x' = \frac{x(c - v)}{c}$$

Since from (1) $x' = ct'$ from (9) we have

$ct' = ct - vt$, or

(11)

$$t' = \frac{t(c-v)}{c}$$

However, equations (10) and (11) differ from the corresponding ones in the Lorentz Transformation. The question is, why?