EXCERPTED FROM

STEPHEN WOLFRAM
A NEW KIND OF SCIENCE

SECTION 9.13

Space, Time and Relativity

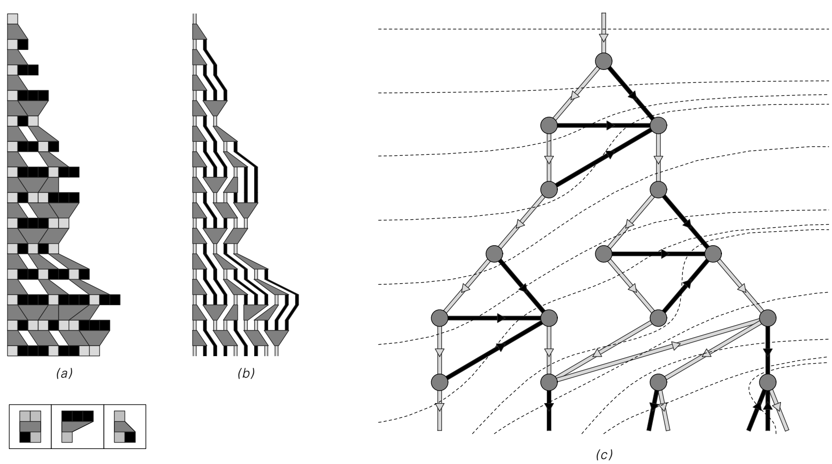
Space, Time and Relativity

Several sections ago I argued that as observers within the universe everything we can observe must at some level be associated purely with the network of causal connections between events in the universe. And in the past few sections I have outlined a series of types of models for how such a causal network might actually get built up.

But how do the properties of causal networks relate to our normal notions of space and time? There turn out to be some slight subtleties—but these seem to be exactly what end up yielding the theory of relativity.

As we saw in earlier sections, if one has an explicit evolution history for a system it is straightforward to deduce a causal network from it. But given only a causal network, what can one say about the evolution history?

The picture below shows an example of how successive steps in a particular evolution history can be recovered from a particular set of slices through the causal network derived from it. But what if one were to choose a different set of slices? In general, the sequence of strings that one would get would not correspond to anything that could arise from the same underlying substitution system.



An example of how the succession of states in an evolution history can be recovered by taking appropriate slices through a causal network. Any consistent choice of such slices will correspond to a possible evolution history—with the same underlying rules, but potentially a different scheme for determining the order in which to apply replacements.

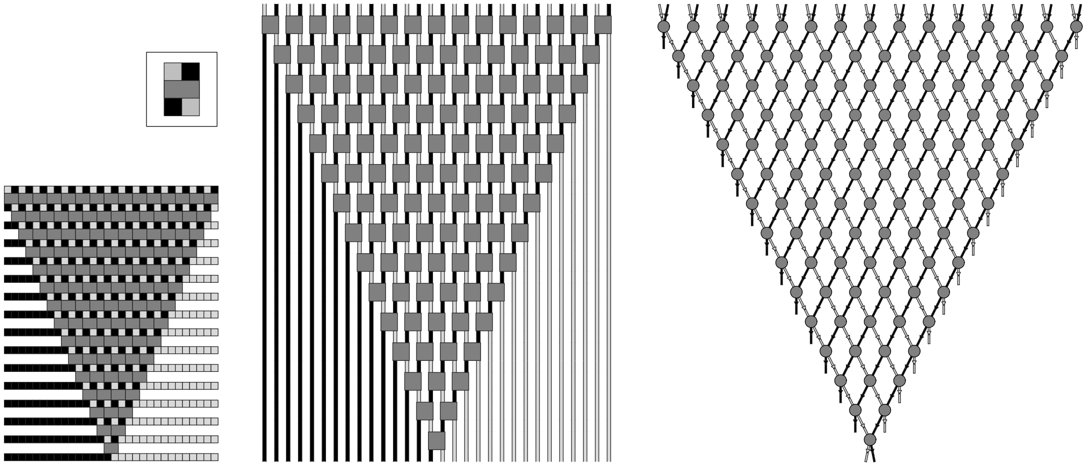
But if one has a system that yields the same causal network independent of the scheme used to apply its underlying rules, then the situation is different. And in this case any slice that consistently divides the causal network into a past and a future must correspond to a possible state of the underlying system—and any non-overlapping sequence of such slices must represent a possible evolution history for the system.

If we could explicitly see the particular underlying evolution history for the system that corresponds to our universe then this would in a sense immediately provide absolute information about space and time in the universe. But if we can observe only the causal network for the universe then our information about space and time must inevitably be deduced indirectly from looking at slices of causal networks.

And indeed only some causal networks even yield a reasonable notion of space at all. For one can think of successive slices through a causal network as corresponding to states at successive moments in time. But for there to be something one can reasonably think of as space one has to be able to identify some background features that stay more or less the same—which means that the causal network must yield consistent similarities between states it generates at successive moments in time.

One might have thought that if one just had an underlying system which did not change on successive steps then this would immediately yield a fixed structure for space. But in fact, without updating events, no causal network at all gets built up. And so a system like the one at the top of the next page is about the simplest that can yield something even vaguely reminiscent of ordinary space.

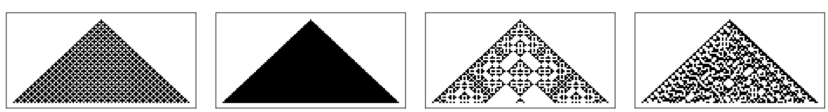
In practice I certainly do not expect that even parts of our universe where nothing much seems to be going on will actually have causal networks as simple as at the top of the next page. And in fact, as I mentioned at the end of the previous section, what I expect instead is that there will always tend to be all sorts of complicated and seemingly random behavior at small scales—though at larger scales this will typically get washed out to yield the kind of consistent average properties that we ordinarily associate with space.



A very simple substitution system whose causal network has slices that can be thought of as corresponding to a highly regular idealization of one-dimensional ordinary space. The rule effectively just sorts elements so that black ones come first, and yields the same causal network regardless of what updating scheme is used.

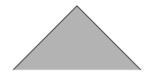
One of the defining features of space as we normally experience it is a certain locality that leads most things that happen at some particular position to be able at first to affect only things very near them.

Such locality is built into the basic structure of systems like cellular automata. For in such systems the underlying rules allow the color of a particular cell to affect only its immediate neighbors at each step. And this has the consequence that effects in such systems can spread only at a limited rate, as manifest for example in a maximum slope for the edges of patterns like those in the pictures below.



Examples of patterns produced by cellular automata, illustrating the fact discussed in Chapter 6 that the edge of each pattern has a maximum slope equal to one cell per step, corresponding to an absolute upper limit on the rate of information transmission—similar to the speed of light in physics.

In physics there also seems to be a maximum speed at which the effects of any event can spread: the speed of light, equal to about 300 million meters per second. And it is common in spacetime physics to draw "light cones" of the kind shown at the right to indicate the region that will be reached by a light signal emitted from a particular position in space at a particular time. So what is the analog of this in a causal network?



Schematic illustration of a light cone in physics. Light emitted at a point in space will normally spread out with time into a cone, whose cross-section is shown schematically here.

The answer is straightforward, for the very definition of a causal network shows that to see how the effects of a particular event spread one just has to follow the successive connections from it in the causal network.

But in the abstract there is no reason that these connections should lead to points that can in any way be viewed as nearby in space. Among the various kinds of underlying systems that I have studied in this book many have no particular locality in their basic rules. But the particular kinds of systems I have discussed for both strings and networks in the past few sections do have a certain locality, in that each individual replacement they make involves only a few nearby elements.

One might choose to consider systems like these just because it seems easier to specify their rules. But their locality also seems important in giving rise to anything that one can reasonably recognize as space.

For without it there will tend to be no particular way to match up corresponding parts in successive slices through the causal networks that are produced. And as a result there will not be the consistency between successive slices necessary to have a stable notion of space.

In the case of substitution systems for strings, locality of underlying replacement rules immediately implies overall locality of effects in the system. For the different elements in the system are always just laid out in a one-dimensional string, with the result that local replacement rules can only ever propagate effects to nearby elements in the string—much like in a one-dimensional cellular automaton.

If one is dealing with an underlying system based on networks, however, then the situation can be somewhat more complicated. For as we discussed several sections ago—and will discuss again in the final sections of this chapter—there will typically be only an approximate correspondence between the structure of the network and the structure of ordinary space. And so for example—as we will discuss later in connection with quantum phenomena—there may sometimes be a kind of thread that connects parts of the network that would not normally be considered nearby in three-dimensional space. And so when clusters of nodes that are nearby with respect to connections on the network get updated, they can potentially propagate effects to what might be considered distant points in space.

Nevertheless, if a network is going to correspond to space as it seems to exist in our universe, such phenomena must not be too important—and in the end there must to a good approximation be the kind of straightforward locality that exists for example in the simple causal network of page 518.

In the next section I will discuss how actual physical entities like particles propagate in systems represented by causal networks. But ultimately the whole point of causal networks is that their connections represent all possible ways that effects propagate. Yet these connections are also what end up defining our notions of space and time in a system. And particularly in a causal network as regular as the one on page 518 one can then immediately view each connection in the causal network as corresponding to an effect propagating a certain distance in space during a certain interval in time.

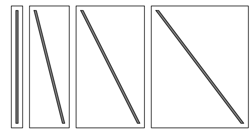
So what about a more complicated causal network? One might imagine that its connections could perhaps represent varying distances in space and varying intervals in time. But there is no independent way to work out distance in space or interval in time beyond looking at the connections in the causal network. So the only thing that ultimately makes sense is to measure space and time taking each connection in the causal network to correspond to an identical elementary distance in space and elementary interval in time.

One may guess that this elementary distance is around 10^{-35} meters, and that the elementary time interval is around 10^{-43} seconds. But whatever these values are, a crucial point is that their ratio must be a fixed speed, and we can identify this with the speed of light. So this means that in a sense every connection in a causal network can be viewed as representing the propagation of an effect at the speed of light.

And with this realization we are now close to being able to see how the kinds of systems I have discussed must almost inevitably succeed in reproducing the fundamental features of relativity theory.

But first we must consider the concept of motion.

To say that one is not moving means that one imagines one is in a sense sampling the same region of space throughout time. But if one is moving—say at a fixed speed—then this means that one imagines that the region of space one is sampling systematically shifts with time, as illustrated schematically in the simple pictures on the right.



Graphical representation in space and time of motion at fixed speeds.

But as we have seen in discussing causal networks, it is in general quite arbitrary how one chooses to match up space at different times. And in fact one can just view different states of motion as corresponding to different such choices: in each case one matches up space so as to treat the point one is at as being the same throughout time.

Motion at a fixed speed is then the simplest case—and the one emphasized in the so-called special theory of relativity. And at least in the context of a highly regular causal network like the one in the picture on page 518 there is a simple interpretation to this: it just corresponds to looking at slices at different angles through the causal network.

Successive parallel slices through the causal network in general correspond to successive states of the underlying system at successive moments in time. But there is nothing that determines in any absolute way the overall angle of these slices in pictures like those on page 518. And the point is that in fact one can interpret slices at different angles as corresponding to motion at different fixed speeds.

If the angle is so great that there are connections going up as well as down between slices, then there will be a problem. But otherwise it will always be the case that regardless of angle, successive slices must correspond to possible evolution histories for the underlying system.

One might have thought that states obtained from slices at different angles would inevitably be consistent only with different sets of underlying rules. But in fact this is not the case, and instead the exact same rules can reproduce slices at all angles. And this is a consequence of the fact that the substitution system on page 518 has the property of causal invariance—so that it gives the same causal network independent of the scheme used to apply its underlying rules.

It is slightly more complicated to represent uniform motion in causal networks that are not as regular as the one on page 518. But whenever there is sufficient uniformity to give a stable structure to space one can still think of something like parallel slices at different angles as representing motion at different fixed speeds.

And the crucial point is that whenever the underlying system is causal invariant the exact same underlying rules will account for what one sees in slices at different angles. And what this means is that in effect the same rules will apply regardless of how fast one is going.

And the remarkable point is then that this is also what seems to happen in physics. For everyday experience—together with all sorts of detailed experiments—strongly support the idea that so long as there are no effects from acceleration or external forces, physical systems work exactly the same regardless of how fast they are moving.

At the outset it might not have seemed conceivable that any system which at some level just applies a fixed program to various underlying elements could successfully capture the phenomenon of motion. For certainly a system like a typical cellular automaton does not—since for example its effective rules for evolution at different angles will usually be quite different. But there are two crucial ideas that make motion work in the kinds of systems I am discussing here. First, that causal networks can represent everything that can be observed. And second, that with causal invariance different slices through a causal network can be produced by the same underlying rules.

Historically, the idea that physical processes should always be independent of overall motion goes back at least three hundred years. And from this idea one expects for example that light should always travel at its usual speed with respect to whatever emitted it. But what if one happens to be moving with respect to this emitter? Will the light then appear to be travelling at a different speed? In the case of sound it would. But what was discovered around the end of the 1800s is that in the case of light it does not. And it was essentially to explain this surprising fact that the special theory of relativity was developed.

In the past, however, there seemed to be no obvious underlying mechanism that could account for the validity of this basic theory. But now it turns out that the kinds of discrete causal network models that I have described almost inevitably end up being able to do this.

And essentially the reason for this is that—as I discussed above—each individual connection in any causal network must almost by definition represent propagation of effects at the speed of light. The overall structure of space that emerges may be complicated, and there may be objects that end up moving at all sorts of speeds. But at least locally the individual connections basically define the speed of light as a fixed maximum rate of propagation of any effect. And the point is that they do this regardless of how fast the source of an effect may be moving.

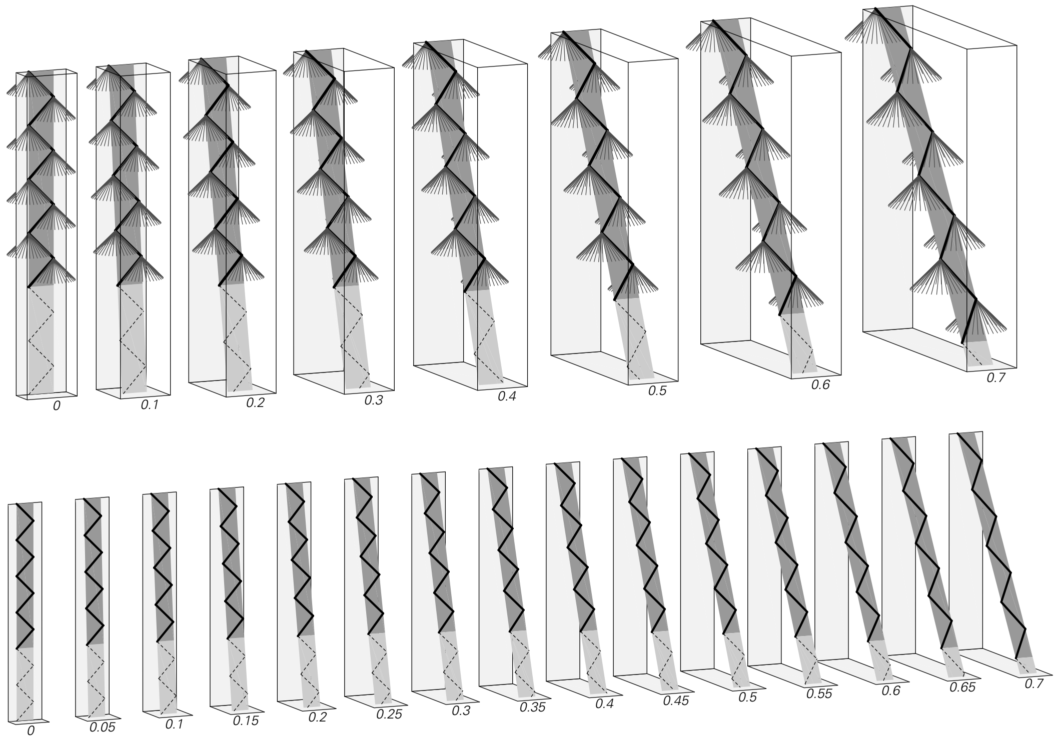
So from this one can use essentially standard arguments to derive all the various phenomena familiar from ordinary relativity theory. A typical example is time dilation, in which a fixed time interval for a system moving at some speed seems to correspond to a longer time interval for a system at rest. The picture on the next page shows schematically how this at first unexpected result arises.

The basic idea is to consider what happens when a system that can act as a simple clock moves at different speeds. At a traditional physics level one can think of the clock as having a photon of light bouncing backwards and forwards between mirrors a fixed distance apart. But more generally one can think of following criss-crossing connections that exist in some fixed fragment of a causal network.

In the picture on the next page time goes down the page. The internal mechanism of the clock is shown as a zig-zag black line—with each sweep of this line corresponding to the passage of one unit of time.

The black line is always assumed to be moving at the speed of light—so that it always lies on the surface of a light cone, as indicated in the top row of pictures. But then in successive pictures the whole clock is taken to move at increasing fractions of the speed of light.

The dark gray region in each picture represents a fixed amount of time for the clock—corresponding to a fixed number of sweeps of the black line. But as the pictures indicate, it is then essentially just a matter of geometry to see that this dark gray region will correspond to progressively larger amounts of time for a system at rest—in just the way predicted by the standard formula of relativistic time dilation.



A simple derivation of the classic phenomenon of relativistic time dilation. The pictures show the behavior of a very simple idealized clock going at different fractions of the speed of light. The clock can be thought of as consisting of a photon of light bouncing backwards and forwards between mirrors a fixed distance apart. (At a more general level in my approach it can also be thought of as a fragment of a causal network.) Time is shown going down the page, so that the photon in the clock traces out a zig-zag path. The fundamental assumption—that in my approach is just a consequence of basic properties of causal networks—is that the photon always goes at the speed of light, so that its path always lies on the surface of light cones like the ones in the top row of pictures. A fixed interval of time for the clock—as indicated by the length of the darker gray regions—corresponds to a progressively longer interval of time at rest. The amount of this time dilation is given by the classic relativistic formula $1/\sqrt{1-v^2/c^2}$, where v/c is the ratio of the speed of the clock to the speed of light. Such time dilation is routinely observed in particle accelerators—and has to be corrected for in GPS satellites. It leads to the so-called twin paradox in which less time will pass for a member of a twin going at high speed in a spacecraft than one staying at rest. The fact that time dilation is a general phenomenon not restricted to something like the simple clock shown relies in my approach on general properties of causal networks. Once the basic assumptions are established, the derivation of time dilation given here is no different in principle from the original one given in 1905, though I believe it is in many ways considerably clearer. Note that it is necessary to consider motion in two dimensions—so that the clock as a whole can be moving perpendicular to the path of the photon inside it. If these were parallel, one would inevitably get not just pure time dilation, but a mixture of it and length contraction.