# Some Correspondences between Spatial and Temporal Representations

Antony Galton
*School of Engineering and Computer Science, University of Exeter, UK*

My starting-point is to trace some connections between some of the classic debates about the representation of temporal information in AI, notably

- Times themselves: instants or intervals?
- What goes on in time: states, events, or other "eventualities"?
- How should the above two types of item be related?

The paradigm of pre-relativistic physics is a good place to start. Here time is represented by an ordered set of instants $(T, <)$, isomorphic to the real number line. States are expressed by means of *fluents*, i.e., functions from $T$ to some value range, for example $s(t)$, the position of an object at time $t$.

In this framework, it is easy to talk about *events* so long as they are instantaneous, e.g., a projectile hits the ground at the instant that $h$, its height above the ground (which has been decreasing), becomes zero. But what about events which take time, such as "the projectile falls from the top of the tower to the ground"? We can't associate this event with a single instant, but must instead invoke a range of instants, or interval, e.g., X moves from A to B over the interval $[t_1, t_2]$. This is fine if we still have instants, since we can define an interval, as here, in terms of instants.

But these instants can give rise to a niggling worry: Does X move from A to B over the *open* interval $(t_1, t_2)$, or over the *closed* interval $[t_1, t_2]$? Each alternative has a supporting argument:

1. Since X is at A at $t_1$ but not thereafter, and X is at B at $t_2$ but not before, so that X is neither at A nor at B during $(t_1, t_2)$, the event must be assigned to the interval $[t_1, t_2]$.
2. Since X is not moving at $t_1$ or at $t_2$, and the event consists of X moving, it must be assigned to the interval $(t_1, t_2)$, throughout which it *is* moving.

This kind of worry led Allen (and Hamblin before him) to look for a way of representing intervals without first introducing instants. As a matter of fact, Allen doesn't actually show us how to *represent* intervals, rather he gives us a set of axioms which our intervals must obey. Let's look at three ways in which we might construct representations of intervals so that they obey Allen's axioms, without mentioning instants.

The first way is to suppose, contrary to the standard representation, that the time sequence is discrete. We may suppose there is a minimum length of interval over which change can occur, in effect viewing time as a discrete sequence of "snapshots" each of which takes up an *atomic interval.* We now define a fluent as a function from atomic intervals to values. Other intervals are constructed as aggregations of consecutive atomic intervals. This model is adequate for many purposes. The major drawback is its inability to handle continuous change, but even this is perhaps not as serious as it sounds, since *uniformly* continuous change can be approximated to an arbitrary degree of precision in a discrete-time model.

A second way of introducing intervals without instants is inspired by the *calendar.* This is in constrast to the classic instant-based model which is inspired by the *clock.* An important difference between clocks and calendars is that whereas clocks are continuous, calendars are discrete. The minute hand moves round continuously (in theory at least), whereas the calendar jumps to a new value at the end of the day.

A calendar identifies times in terms of a *nest of cycles.* We have the sequence of years, each of which is subdivided into a sequence of months, each of which is in turn divided into a sequence of days. And although this is not normally regarded as the province of the calendar, we can continue the subdivision, of days into hours (e.g., the tenth hour—extending from 9 a.m. to 10 a.m.—on 24th September 1998), hours into minutes, minutes into seconds, seconds into milliseconds, milliseconds into microseconds, or whatever system you want to use. This gives us a plentiful supply of intervals, which in principle can be of arbitrarily small length.

This system is fine if our prototypical way of locating an event in time is to say "E occurs within the interval I", for example, "Antony Galton flew from London to Venice on 23rd September 1998". But Allen wants his notation $Occur(E, I)$ to mean that $E$ occurs *on* the interval $I$, where $I$ is the exact time of occurrence of $E$, beginning when $E$ begins and ending when $E$ ends, with no time left over on either side. For this purpose we need to create new intervals by aggregating sequences of sufficiently fine-grained consecutive "basic" intervals into intervals that can be used to locate events with arbitrary accuracy.

Total accuracy is not always possible, even in principle, unless one is prepared to aggregate infinite sets of basic intervals. Consider the event: X moves at 1 m/s along the diagonal of a square measuring 1 metre square. The interval on which this event occurs is not any finite aggregate of decimal subdivisions of a second: if it starts at the beginning of 1998, then it occupies the first second of 1998, plus the first four tenths of the second immediately following, plus the first hundredth of a second immediately following those, and so on, spelling out the non-terminating decimal expansion of $\sqrt{2}$.

Well, this is a frivolous example, since it presupposes arbitrary accuracy of measurement, which we all know to be impossible. Still, it does highlight the fact that assigning intervals to events can be awkward if the intervals have to be expressed as aggregates of finite sequences from some discrete system of calendrical cycles laid down in advance.

This brings us to our third way of introducing intervals, which is simply to *define* intervals in terms of the events that occur on them. To get Allen's axioms to hold, we simply recast them as axioms for events rather than for intervals. Note that these are not quite the same; for example, Allen has an axiom which says that if I meets both J and J′, *and* both J and J′ meet K, then J must be the same interval as J′. Well, if I, J, J′, and K are events rather than intervals, then J and J′ needn't be the same event, but they must be strictly simultaneous (i.e., beginning and ending together).

Here we have begun to sketch out a system in which the primitive givens are events, together with some temporal relations defined over them—minimally, simultaneity and succession. Time is then defined in terms of these: an interval is an equivalence class of events under the relation of strict simultaneity.

What we have seen here is a contrast between two different approaches to the representation of temporal information. We might call these the *absolute* and *relational* approaches.

In the absolute approach, we start by laying down a fixed temporal framework, consisting of instants, intervals, or both, and then populate this framework with eventualities (events or fluents). Here, temporal relations between eventualities are derived from temporal relations between times: e.g., $E_1$ precedes $E_2$ if the time associated with $E_1$ precedes the time associated with $E_2$.

In the relational approach, we start with a set of events and temporal relations over these, in terms of which we go on to define times as derived entities. Temporal relations between times are then derived from temporal relations between events: e.g., interval $I_1$ precedes interval $I_2$, where $I_1$ and $I_2$ are equivalence classes of events, if any event in class $I_1$ precedes any event in class $I_2$. For this to work, we require appropriate axioms relating simultaneity and succession, e.g., if I is simultaneous with J, which precedes K, then I precedes K.

The relational approach is best adapted to talking about events rather than about fluents. Consider the fluent $s$ giving the position of an object at each instant. How could this be expressed in a relational system? We have to pick out some events which can enter into temporal relations with other events. An event that might be used as part of the characterisation of fluent $s$ is "The object moves from position $s_1$ to position $s_2$". There may be any number of occurrences of this event, but each occurrence enjoys definite temporal relations with each occurrence of the infinite number of other such events, and some subset of the totality of these relations could serve, in principle, to characterise the fluent $s$ exactly. The problem is that there is no canonical way of choosing such a subset: the whole procedure is both arbitrary and cumbersome.

Thus the absolute system is more natural for describing some types of information, e.g., continuously-varying fluents, while the relational system is more natural for other types, e.g., discrete events. Information presented in one way can, in principle, be converted into the other mode of presentation, but at considerable cost in "naturalness".

This observation still holds good, suitably modified, if we renounce the purely relational aspect of our relational system. In presenting the relational system *as* relational, we are assuming that initially events are given without any concept of times, which are defined later. A compromise between this and an absolute system can be described as follows. We have a temporal framework specified in advance, as in the absolute system. We then introduce events, each event being associated with a certain time in the temporal framework. This is still significantly different from the "classical" absolute approach where we have fluents, and each instant in the temporal framework is associated with a value of the fluent. The difference is simply that, *for events, we map occurrences onto times, whereas for fluents, we map times onto values.*

We have here both an *event-based system* and a *fluent-based system.* Since we will want to talk about both events and fluents (including states), we shall need to have both systems in operation. They share a common temporal framework, but the nature of their relationship to this framework differs. With an event, the natural question is "When did it happen?", answered by means of the function which maps event-occurrences to times, whereas with a fluent, the natural question is "What it its value at time $t$?", which is answered by means of the function which maps times to fluent-values.

* * *

At this point I am going to shift the focus rather abruptly from temporal information to spatial information, in particular the kinds of information that one typically considers in geographical information systems (GIS). Here it is customary to draw a distinction between two different ways of presenting the data, known as the *raster* method and the *vector* method.

In the raster method, a spatial framework, consisting of a set of *locations*, is assumed, and a *field* is defined over this framework by assigning a value to each location. At an idealised mathematical level the spatial framework might be conceived of as the consisting of the points of a continuous space, e.g., the surface of a sphere (representing the surface of the earth), but in practice this will be discretised to a grid of minimal cells (e.g., 10 km squares).

The value ranges associated with the field may be diverse, from the purely qualitative (e.g., different vegetation types) to the fully quantitative (e.g., elevation, rainfall). For example, to each point (or raster cell) on the earth's surface we might assign a number representing its elevation above or below sea-level (thus defining a quantitative field), and also a value from the set {"land", "fresh water", "salt water"}, defining a qualitative field.

In the vector method, the data consists of a collection of *objects* or *entities*, each of which has associated with it a set of attributes, including spatial attributes such as size, shape and position. The spatial attributes are typically given in the form of vectors specifying points in space. An object may be a point object, specified by means of a single vector, a line object, specified by means of a sequence of vectors, or an area object, or polygon, specified by means of a closed sequence of vectors defining its boundary.

Examples of such objects are towns (which are represented as point objects on a small scale, or area objects on a larger scale), rivers and roads (line objects), and forests, lakes, islands, and administrative districts (area objects).

Raster and vector representations are interconvertible, typically with some loss of information. For example, given a portion of the earth's surface for which each point is allocated one of the attributes "land" and "sea", the set of "land" points may constitute an island, which can be represented as an area object, a polygon containing all and only the "land" points in its interior. Or again, suppose a country is divided up into districts, as Italy is divided into Veneto, Lombardia, Toscana, Umbria, Lazio, Puglia, etc, each of these regions being an area object. But each point in Italy belongs to one or other of these objects, so a field can be defined which maps each point onto a value in the set {Veneto, Lombardia, Toscana, Umbria, Lazio, Puglia, ...}.

There has been a long-standing debate within the GIS community regarding the relative merits of the raster and vector approaches. Nowadays, it is generally accepted that both are needed and that, going beyond this, we require a flexibility of approach which does not sharply dichotomise all our methods into one of the other of these two types.

Now I think it should be obvious from what I have said here that the field-based (raster) and object-based (vector) approaches to spatial data are the exact analogues of the fluent-based and event-based approaches to temporal data. Consider the correspondences:

> A fluent is a mapping from times to values.
> A field is a mapping from places to values.
>
> An event is a temporal entity occurring at a (possibly extended) time.
> A (spatial) object is a spatial entity existing at a (possibly extended) place.

And just as we noted that the natural thing to ask of an event is when it happened, whereas the natural thing to ask of a fluent is its value at a given time, so in GIS the distinction between the vector and raster approaches is often characterised by noting that whereas the former is particularly well adapted to answering questions of the form “Where is X?”, the latter is more appropriate to questions of the form “What is there at such-and-such a location?”.

Please note that we are here describing these things at the level of a *conceptual* rather than a *computational* model. By the object-based approach I do not necessarily mean the use of object-oriented programming techniques, although there is certainly an affinity between these things. But a field-based model does not become an object-based model just because it is implemented using an object-oriented language; a field can be implemented as an object in the programming context without it thereby become an object at the level of the original conceptual model. The kind of object, in the programming sense, that a field can be implemented as is still very different from the kind of object that would be used to implement the objects of an object-based model. And equally, one could implement the objects of the conceptual model without using object-oriented programming techniques.

* * *

The close correspondence that we have noted between the spatial and temporal domains ties in with an observation that goes back many years and that has been drawn attention to by a number of different authors in linguistics and the philosophy of language, and this brings me to my second abrupt shift of focus. I refer to the analogy between, on the one hand, the *aspectual* distinction between imperfective and perfective verbs, and on the other hand, the distinction between mass nouns and count nouns.

The first of these two distinctions can be briefly explained as follows.

Some verbs, typically denoting states and activities, are regarded as inherently imperfective because the situations they refer to have a certain homogeneity (in that a period over which such a situation holds can be divided up into subperiods over each of which that situation still holds), and, related to this, they lack any notion of completion, meaning that there is no distinction, for these verbs, between finishing and merely stopping. For example, I must be sleeping during any subperiod of a period during which I am sleeping, and when I wake up, I thereby stop sleeping, but I do not in any sense finish sleeping. Similar remarks apply to an activity such as running: if I run for an hour then I run for the first half hour and the second half hour; and if I stop running then I have run, nothing else is needed to ensure “completion”.

By contrast, other verbs (or more accurately, verb phrases) are inherently perfective, because they refer to events or occurrences which are unitary (i.e., they apply to an interval as a whole, without applying to its proper subintervals) and they have a built-in notion of completion, allowing us to distinguish between their finishing and their merely stopping. These verbs are often obtained from the imperfective ones by adding a limiting qualification, e.g., running a mile instead of just running. If I run a mile over a certain interval, then I do not run a mile (but some lesser distance) over any of its proper subintervals; and I can be running a mile, and then stop doing so before I have completed the mile, say at the half-mile point,

We can broadly characterise the distinction here as one between *event-verbs* (perfective) and *fluent-verbs* (imperfective), thus bringing it into line with our earlier distinction. Running is a Boolean-valued fluent in that at each moment I either am or am not running, so we have a function from times to truth-values. Running a mile is different: there are certain intervals over which I run a mile and each occurrence of the generic event-type “I run a mile” is associated with a particular unique interval. (Note however that “I am running a mile”, with the progressive form of the verb, is more like a fluent: the function of the progressive in English is precisely to construct a fluent out of an event.)

This distinction between so-called aspectual characters has been the subject of intensive investigation for many years, both at a rather generalised philosophical level (see my book *The Logic of Aspect*, for example) and, more concretely, with reference to the verb-systems of different particular languages. And scattered amongst this literature one frequently comes across references to an analogy with another distinction, applying to nouns and noun phrases rather than verbs and verb phrases, namely, the distinction between mass terms and count terms.

Count nouns are nouns which paradigmatically refer to discrete objects, e.g., “book”, “table”, “man”, and “star”. Grammatically, the general mark of a count noun is that it can occur in the singular and plural, and more specifically, in English, count nouns co-occur with adjectives such as “many”, “few” and the cardinal numbers.

Mass nouns, by contrast, refer to types of matter or stuff, with no implication as to discrete chunking, e.g., "water", "wool", "mud", and "salt". Grammatically, mass nouns are characterised by their not taking the plural and, in English, co-occurring with adjectives such as "much" and "little", as well as with measure phrases such as "a litre of" or " a kilogram of".

Many nouns can be used as both mass and count nouns, typically with different though related senses. Quine has a nice example of this: "Mary had a little lamb". If we are talking about Mary's pets, then "lamb" here is a count noun; but if we are talking about her Sunday dinner, it is a mass noun.

Mass nouns, like imperfective verbs, are homogeneous: any part of a sample of water/wool/mud/salt is again water/wool/mud/salt. Count nouns, like perfective verbs, are unitary: no proper part of a book/table/man/star is itself a book/table/man/star. This has led some linguists to postulate an underlying abstract distinction which is neutral between nouns, verbs, and perhaps other parts of speech as well; the distinction materialises as the mass/count distinction when it arises in conjunction with nouns and noun phrases, but as the imperfective/perfective distinction when it arises in conjunction with verbs and verb phrases.

We can tie in the mass/count distinction with the field-based and object-based approaches to GIS. Consider the idea of "forest", for example. In English, "forest" can be either a mass noun or a count noun. We can say "the northern part of the country is mostly covered with forest" (mass noun), and "there are three forests of area in excess of 10,000 sq km to the north of the city" (count). How might these statements be represented in a GIS? For the first statement we might expect to see a Vegetation field, mapping locations onto a set of vegetation-types such as forest, grassland, desert. We would find that a high percentage of the locations in the northern part of the country in question have forest as their vegetation type. For the second statement, we would expect to find three (and only three) distinct area objects, furnished with spatial coordinates from which one could infer that they are situated to the north of the city and each has an area in excess of 10,000 km.

*　*　*

I have discussed here a range of different approaches to representing temporal and spatial data. I have not laboured the details of these approaches, preferring rather to present a *bird's-eye view* of the field, paying particular attention to the close correspondences which exist between contrasting pairs of approaches in both the temporal and spatial domains: fluent-based and event-based methods in the former, and field-based and object-based in the latter. I have further drawn some connections with another body of work, in linguistics and philosophy, showing how the distinctions between imperfective and perfective verbs, and between mass and count nouns, are aligned with the other distinctions.

What can be learnt from these correspondences? One hope might be that one can construct a theoretical framework at a high level of generality that is applicable to both the temporal and spatial domains and which can form the basis of a uniform scheme for implementing both temporal and spatial data structures. Of particular interest is the problem of handling both spatial and temporal data in the same system. One way of approaching this is to bring together an adequate scheme for handling temporal data with an adequate scheme for handling spatial data, and somehow tying them together so that one has an adequate scheme for handling data involving both space and time (e.g., involving change in space). Since both types of data can be approached in two ways—fluents versus events, and fields versus objects—the possibilities for tying them together are manifold, and the implications of, for example, mixing a fluent-based temporal scheme with an object-based spatial one, have not been systematically explored in detail.

Another possibility is to develop an integrated spatio-temporal system with an underlying four-dimensional framework of "locations" populated by four-dimensional "entities". Within such a system, the field-based versus object-based dichotomy again arises, and again this is something that has not been thoroughly investigated.

Finally, what of the linguistic dimension? Perhaps we can exploit the parallels we have noted there to produce, for example, a formal semantics for natural language that directly references the structures in a computational scheme for handling spatio-temporal data, enabling us perhaps to construct a workable natural-language interface to a spatio-temporal database or geographical information system.

This talk, then, should be seen as more than just a small contribution to the history of ideas (which it is), but also as a call for action on the part of researchers in all these areas to create schemes of high generality that will embody the common threads that we have discerned running through them all.