Towards a Qualitative Theory of Movement

Antony Galton
Department of Computer Science
University of Exeter
Exeter EX4 4PT, UK
email: antony@dcs.exeter.ac.uk

Abstract

The phenomenon of movement arises whenever the same object occupies different positions in space at different times. Therefore a theory of movement must contain theories of time, space, objects, and position. We provide a theoretical basis for describing movement events in terms of the conditions for their occurrence, which refer to the holding or not holding of various positional fluents at different times. For this we need to bring together a formal model of time with a formal model of space. By attending closely to the constraints imposed by continuity on the temporal behaviour of different fluents we develop theory of *dominance*, which enables us to generate *ab initio* the perturbation relation on the full set of positional relations.

1 Introduction

The phenomenon of movement arises whenever the same object occupies different positions in space at different times. This bare definition already suggests the main ingredients of a theory of movement:

1. *A theory of time*, comprising (a) a theory of *times*, i.e., a fundamental set of temporal entities, usually either instants or intervals, which act as loci in the temporal dimension for things to happen in, and (b) a theory of *temporal order* by which the fundamental set of times is endowed with an ordering relation, representing temporal sequence.

2. *A theory of space*, analogous to the theory of time but with points or regions as the fundamental entities, and a more complex theory of spatial ordering.

3. *A theory of objects*, which allows objects to be rigid or non-rigid (the latter being capable of a wider variety of modes of motion than the former), and to have parts, and which can include objects which are not discrete individuals (consider the water in a river—which certainly moves), or which are not even concrete substances at all (e.g., holes and shadows, both capable of movement).

4. *A theory of position*, which brings together the theory of objects with the theory of space, the idea being that, at any one time, each object occupies a certain part of space, which is called its position at that time. The position of an object can be specified as the total region of space occupied by it at a time. An object's position will thus be a region of space precisely congruent, in the geometrical sense, to the body itself.

Our aim is to provide a framework for formalising our "common-sense" knowledge of the world. This is a different enterprise from that of physics, which seeks to go beyond our common-sense view, correcting it where it is in error, and providing a unified explanatory framework. But we should not ignore the fact that a well-developed physico-mathematical theory of motion exists; and ideally our common-sense view should be reconcilable with the physico-mathematical view. A common-sense theory is typically qualitative in nature, whereas the physico-mathematical picture is quantitative. We shall be particularly interested in examining how phenomena such as continuity, which find a natural expression in the quantitative theory, can be expressed when we move over to a qualitative theory. Throughout, it should be borne in mind that all theories, whether qualitative or quantitative, naive or scientific, are idealisations and as such cannot be regarded as absolutely true to the complete underlying reality.

2 Theory of Time

We shall use an instant-based model $\mathcal{T} = (T, <)$, which posits a set T of entities called *instants* to be the fundamental set of temporal loci. They are endowed with a relation $<$ of *temporal succession*. We use abbreviations as follows:

$$\begin{aligned} t \leq u \quad &\stackrel{\Delta}{=} \quad t < u \vee t = u \\ t < u < v \quad &\stackrel{\Delta}{=} \quad t < u \wedge u < v \end{aligned}$$

Properties typically ascribed to this relation (cf. (van Benthem 1983)) are as follows (we use the capitalised parts of the names as labels for the axioms):

1. IRReflexivity. No time precedes itself:

$$\forall t \neg (t < t).$$

2. TRAnsitivity. If time t precedes time u which in turn precedes time v, then t also precedes v:

$$\forall t, u, v (t{<}u{<}v \rightarrow t{<}v).$$

3. LINearity. Of any two distinct times, one precedes the other:

$$\forall t, u (t \neq u \rightarrow t < u \vee u < t).$$

4. UNBoundedness. Every time has a time preceding it, and a time which it precedes (so there are no first and last times):

$$\forall t \exists u, v (u < t < v).$$

5. DENsity. Between any two times there is a third (and hence, given IRR and TRANS, infinitely many):

$$\forall t, u(t<u \rightarrow \exists v(t<v<u)).$$

An *interval* is defined by specifying, for each instant, whether it precedes, begins, divides, ends, or follows the interval. It is sufficient to specify just two instants, namely the unique instants at which the interval begins and ends. We could thus define an interval as an ordered pair $\langle t, u\rangle$ of instants, where $t < u$. We call this the *pair model* of an interval. The instants which precede $\langle t, u\rangle$ are precisely the instants which precede t, and the instants which follow $\langle t, u\rangle$ are the ones which follow u.

If $i = \langle t, u\rangle$, we write $Beg(i) = t$ and $End(i) = u$. This notation allows us to refer to the instants marking the beginning and end of any interval we can refer to. We write $Lim(t, i)$ whenever either $t = Beg(i)$ or $t = End(i)$, and say that t *limits* i in this case. We write $t \sqsubset i$, and say that t *divides* i, as an abbreviation for $Beg(i) < t < End(i)$. The motivation for this term is simply that t divides i into two contiguous subintervals $\langle Beg(i), t\rangle$ and $\langle t, End(i)\rangle$.

An interval is often identified with a range of instants, namely all those instants which divide it. For the interval $\langle t, u\rangle$ these are the instants v such that $t < v < u$. The instants in the range are thus members of the set $\{v \mid t < v < u\}$, and it is customary to *identify* the interval with this set. We call this the 'set model' of an interval. Although it is widely used in mathematical modelling, the set model of an interval has nothing to offer over and above the pair model. In particular, the distinction between open and closed intervals which is suggested as a natural extension of the set model does not seem to have any application either to physical time or to any of our common-sense conceptions of time.

There are three distinct temporal successions relation on intervals, and any of them can be taken as fundamental. Their definitions are

- *immediate succession*, in which the first interval 'meets' the second at an instant, without any intervening interval:

  $$\langle t, u\rangle | \langle v, w\rangle \quad \stackrel{\Delta}{=} \quad u = v.$$

  The instant at which the intervals meet in this case is u $(= v)$.

- *delayed succession*, in which the first interval is separated from the second by an intervening interval:

  $$\langle t, u\rangle < \langle v, w\rangle \quad \stackrel{\Delta}{=} \quad u < v.$$

  The intervening interval in this case is $\langle u, v\rangle$.

- *general sucession*, which covers the preceding two cases:

  $$\langle t, u\rangle \lhd \langle v, w\rangle \quad \stackrel{\Delta}{=} \quad u \leq v.$$

General succession is the disjunction of immediate and delayed succession, which are themselves mutually incompatible. We write things like $i|j|k$ and $i < j < k$ with the obvious meanings. We write $i][j$ to denote the instant at which i meets j (so if $i = \langle t, u\rangle$ and $j = \langle u, v\rangle$ then $i][j = u$). Note the 'polymorphic' character of the symbol '$<$', which we use both as a relation on instants and as a relation on intervals; it is natural to think of these as the 'same' relation (since with instants, succession is always delayed succesion). Note, however, that our axioms for $<$ are intended to apply only to the relation on instants (in particular, LIN does not hold when the terms of the relation are intervals).

Using our axioms for $<$ we can prove a number of important properties of $|$, as follows (the proofs are easy and are omitted here—but we indicate which axioms we need to use; where no axioms are cited, nothing more is needed than the definitions of $|$ and $<$). In M2, '$\oplus$' is the exclusive 'or' connective.

(M1)	$\forall i, j, k, l(i\|k \wedge i\|l \wedge j\|k \rightarrow j\|l)$	[No axioms]
(M2)	$\forall i, j, k, l(i\|k \wedge j\|l \rightarrow i < l \oplus i\|l \oplus j < k)$	[LIN]
(M3)	$\forall i \exists j, k(j\|i\|k)$	[UNB]
(M4)	$\forall i, j, k, l(i\|j\|l \wedge i\|k\|l \rightarrow j = k)$	[No axioms]
(M5)	$\forall i, j[i\|j \rightarrow \exists k \forall l[(l\|k \leftrightarrow l\|i) \wedge (k\|l \leftrightarrow j\|l)]]$	[TRA]
(M6)	$\forall i \exists j, k[j\|k \wedge \forall l[(l\|i \leftrightarrow l\|j) \wedge (i\|l \leftrightarrow k\|l)]]$	[DEN]

Note that M4 and M5 together imply that if i meets j then there is a unique interval which meets whatever j meets and is met by whatever i is met by (existence is given by M5, uniqueness by M4). We denote this interval $i + j$; it is the interval which begins when i begins and ends when j ends, and hence spans the entire time taken up by i and j together.

We can define further relations on intervals as follows. We use infix notation after the fashion of (Allen 1983):

$$
\begin{aligned}
iOj &\triangleq \exists i', j', k(i = i' + k \wedge j = k + j') \\
iSj &\triangleq \exists k[j = i + k] \\
iDj &\triangleq \exists k, l[j = k + i + l] \\
iFj &\triangleq \exists k[j = k + i] \\
i \sqsubset j &\triangleq iSj \vee iDj \vee iFj \\
i \sqsubseteq j &\triangleq i \sqsubset j \vee i = j
\end{aligned}
$$

(O, S, D, F are read as 'overlaps', 'starts', 'is during', and 'finishes', respectively.) The relations $\sqsubseteq$ and $\sqsubset$ are analogous to the set-theoretic relations of subset ($\subseteq$) and proper subset ($\subset$), and indeed could be defined to be identical to them if the set model of an interval were to be adopted.

Following Allen, it is often regarded as more satisfactory to base one's temporal model on intervals as the fundamental temporal units rather than instants. Allen and Hayes (1985) do this, using equivalents of (M1)–(M6) as their axioms for immediate succession. They also show how instants can be defined in their system; the resulting system satisfies our axioms for instants. It follows that it does not matter whether we base our temporal model on instants or intervals. By choosing appropriate definitions we will end up with equivalent systems starting from either choice. It is certainly convenient to be able to use both instant and interval notations!

The model of time presented here is dense, because of axiom M6. The obvious mathematical model for dense time, satisfying all our axioms, is to represent instants by real numbers, with temporal succession represented by the 'less than' relation. We could either choose the set of all real numbers to represent T, or any dense subset of them, such as the rational numbers. Dense time implies that there is no lower limit to the length of an interval, and while this is a satisfactory idealisation for many purposes, sometimes there are good reasons not to accept it. For example, in a context in which intervals are only known through observation and measurement, there is an effective lower limit to the length of an interval, corresponding to our chronometrical resolving power; or again, one may be concerned with physical processes which always require a certain minimum duration in which to occur.

For these reasons, an alternative *discrete* model of time is often preferred: mathematically, this is tantamount to representing instants as integers rather than real numbers. In our axiomatisation, it is necessary to replace the axiom DEN by

6. DISCreteness. If time t precedes time u then there is an earliest time v which t precedes and a latest time v' which precedes u:

$$\forall t, u(t < u \rightarrow \exists v \forall w(t < w \leftrightarrow v \leq w) \wedge \exists v' \forall w(w < u \leftrightarrow w \leq v)).$$

Note how we say that v is the earliest time which t precedes: an arbitrary time w is preceded by t if and only if it is either equal to or preceded by v, so that the only times which t precedes are v and anything which v precedes; and analogously for the latest time which precedes t.

In discrete time, the elements of T are *atomic intervals*, or *moments*. A general interval in discrete time is the concatenation of one or more consecutive moments, a natural measure of its duration being the number of moments involved. This is quite different from dense time, where it makes no sense to speak of 'consecutive instants', and duration has to be introduced as an additional primitive, not derivable from the temporal order alone. As will be seen below, we still need the notion of 'instant' in discrete time: we shall need to speak of the instant at which two consecutive moments meet. The reason for this will become apparent when we consider the occurrence conditions for an instantaneous event in the next section.

In the rest of this paper we shall mostly confine our attention to dense rather than discrete time. None the less, much of what we say about movement in the next section can be adapted fairly straightforwardly to the latter case. On the other hand, the later material on continuity and the theory of dominance applies specifically to dense time, and does not make much sense in the discrete case.

3 Movements and their Occurrence Conditions

We use the RCC-8 system (Randell, Cui and Cohn 1992) for specifying relations between regions. There are eight basic relations, as follows:

DC	A is disconnected from B
EC	A is externally connected to B
PO	A partially overlaps B
EQ	A is equal to B
TPP	A is a tangential proper part of B
NTPP	A is a non-tangential proper part of B
TPPI	A has B as a tangential proper part
NTPPI	A has B as a non-tangential proper part

These eight relations correspond closely to the eight relations determined by Egenhofer's 4-intersection method (Egenhofer 1991). Regions will be mainly of interest to us as possible positions for movable bodies. The position of a body can be given, with greater or less precision, by the RCC-8 relation which it bears to some known region, such as the position of another body. To be able to talk about motion, we need only relativise this to time.

We take as our fundamental notion for the analysis of change the idea of a *fluent*. A fluent can take different values at different times. If f is a fluent and a is a value it can take, then $f = a$ is a proposition that can be true or false at different times, in other words a *Boolean fluent*, or *state*. Likewise, if f_1 and f_2 are fluents, and R is a relation which may hold between values that they can take, then $R(f_1, f_2)$ is also a state.

We write $Holds\text{-}at(S, t)$ to indicate that state S holds at instant t, and $Holds(S, i)$ to indicate that S holds throughout the interval i (Allen 1984, Galton 1990). These two notations are connected by the rule

$$Holds(S, i) \leftrightarrow \forall t \sqsubseteq i Holds\text{-}at(S, t),$$

which says that a state holds throughout an interval if and only if it holds at every instant which divides the interval. This could be taken as a definition of $Holds$ in terms of $Holds\text{-}at$, if desired. An immediate consequence is Allen's rule

$$Holds(S, i) \wedge i \sqsubseteq j \rightarrow Holds(S, j),$$

which says that a state holds throughout every subinterval of any interval throughout which it holds.

We write $S \sqcap S'$ to refer to the state which holds when and only when both S and S' hold (*state-conjunction*), and $-S$ to refer to the state which holds when and only when S fails to hold (*state-negation*). Formally, they obey the rules

$$\begin{aligned} Holds\text{-}at(S \sqcap S', t) &\leftrightarrow Holds\text{-}at(S, t) \wedge Holds\text{-}at(S', t), \\ Holds\text{-}at(-S, t) &\leftrightarrow \neg Holds\text{-}at(S, t). \end{aligned}$$

We write $pos(a)$ to denote the position of body a. This is a region having exactly the same shape and size as a, so that a can fit into it with no space left over. Since the position of a can change over time, $pos(a)$ is a fluent. We write

$$Holds(R(pos(a), r), i)$$

to indicate that throughout interval i, the position of a bears the RCC relation R to region r.

To handle movement we need a formalism for referring to events, since a movement is an event, not a state. We write $Occurs(e, i)$ to indicate that an event of type e occurs over interval i. Allen's rule for $Occurs$ is

$$Occurs(e, i) \wedge j \sqsubset i \rightarrow \neg Occurs(e, j),$$

which says that an event does not occur over any proper subinterval of an interval over which it occurs. (So events can be described as *unitary*, in contrast to states, which are *homogeneous*.) This rule can be regarded as a constraint on the allowable event-types. We shall also have cause to talk about *instantaneous events*. For these we write $Occurs\text{-}at(e, t)$ to indicate that an event of type e occurs at the instant t (see Galton (1994) for a detailed treatment of instantaneous events).

Our paradigm for analysing movement will be to specify a movement event e in terms of its *occurrence conditions*, that is in terms of a formula of one of the forms

$$\begin{aligned} Occurs(e, i) &\stackrel{\Delta}{=} \cdots \\ Occurs\text{-}at(e, t) &\stackrel{\Delta}{=} \cdots \end{aligned}$$

where the right-hand side is a formula not containing e (Galton 1993, Galton 1994). We shall consider a number of examples.

Suppose we wish to define what it is for a to *move* from position r_1 to position r_2 over the interval i. A natural first attempt might be to stipulate that a must be at r_1 throughout some interval j which meets i, and at r_2 throughout some interval which i meets. In order to ensure that the event occurs over the whole interval i, and not some proper subinterval (in accordance with Allen's rule for $Occurs$), we should add that a is not at either r_1 or r_2 at any time during i itself. This gives us the definition:

$$\begin{aligned} &Occurs(move(a, r_1, r_2), i) \stackrel{\Delta}{=} \\ &\quad \exists j, k(j|i|k \wedge Holds(pos(a) = r_1, j) \wedge Holds(pos(a) = r_2, k)) \wedge \\ &\quad Holds(pos(a) \neq r_1, i) \wedge Holds(pos(a) \neq r_2, i). \end{aligned}$$

This definition is adequate if we do not allow states to be said to hold at instants, for example if our model of time is discrete; but it will not do in general. For suppose a moves from position r_0 to r_3, passing through positions r_1 and r_2, in that order, but without stopping at either of them. Then there is no interval throughout which a is at either r_1 or r_2, yet a still moves from r_1 to r_2 over the interval between the times at which it is at these positions. We must replace our definition of $move$ by:

$$\begin{aligned} &Occurs(move(a, r_1, r_2), i) \stackrel{\Delta}{=} \\ &\quad Holds\text{-}at(pos(a) = r_1, Beg(i)) \wedge Holds\text{-}at(pos(a) = r_2, End(i)) \wedge \\ &\quad Holds(pos(a) \neq r_1, i) \wedge Holds(pos(a) \neq r_2, i). \end{aligned}$$

Note that this definition subsumes the previous one, since if a is at position r_1 over an interval meeting i, then by continuity it must be at r_1 at the beginning of i, and likewise with r_2 at the end.

Suppose next that we wish to say that a *enters* region r over interval i. A natural first attempt would be to postulate that a must be just outside (EC) r throughout some interval j which meets i, and just inside (TPP) r throughout some interval k which i meets; during i itself, a must be partly inside and partly outside (PO) r:

$$\begin{aligned}&Occurs(enter(a,r),i) \quad \stackrel{\Delta}{=} \\ &\quad \exists j,k(j|i|k \wedge Holds(EC(pos(a),r),j)) \wedge \\ &\quad Holds(PO(pos(a),r),i) \wedge Holds(TPP(pos(a),r),k)).\end{aligned}$$

As before, we can argue that a does not need to be EC or TPP to r for more than an instant: consider the case where a approaches r from a distance and enters it without pausing in either the EC or the TPP positions. A more general definition is therefore:

$$\begin{aligned}&Occurs(enter(a,r),i) \quad \stackrel{\Delta}{=} \\ &\quad Holds\text{-}at(EC(pos(a),r),Beg(i)) \wedge \\ &\quad Holds(PO(pos(a),r),i) \wedge Holds\text{-}at(TPP(pos(a),r),End(i))).\end{aligned}$$

Suppose finally we wish to characterise the event of two objects' coming into *contact*; this is an instantaneous event. It can happen at the meeting point of two intervals i, j, such that a is separated from b throughout i and touching b throughout j:

$$Holds(DC(pos(a),pos(b)),i) \wedge Holds(EC(pos(a),pos(b)),j).$$

The event itself occurs at the instant $i][j$. However, this does not cover the case where a moves towards b and then enters it. Then a first makes contact with b at $i][j$, where

$$Holds(DC(pos(a),pos(b)),i) \wedge Holds(PO(pos(a),pos(b)),j)).$$

Here we can infer, by continuity, that a is EC to region r at the instant t. Even this does not cover the case where the objects move apart as soon as they have touched. If the moment of touching is $i][j$, then we have

$$Holds(DC(pos(a),pos(b)),i) \wedge Holds(DC(pos(a),pos(b)),j)$$

but this does not tell us that they ever touched. Hence we have to bring in explicit reference to the state of affairs holding at the instant $i][j$ itself. Our fully general definition, subsuming the others, will therefore be

$$\begin{aligned}&Occurs\text{-}at(connect(a,r),t) \quad \stackrel{\Delta}{=} \\ &\quad \exists i(t = End(i) \wedge Holds(DC(pos(a),r),i) \wedge Holds\text{-}at(EC(pos(a),r),t))\end{aligned}$$

This says that a is touching b at the instant which ends an interval throughout which a is separated from b: that is the instant at which a makes contact with b. Nothing need be said about what state holds after that instant.

Note that in discrete time, the same analysis will apply, but it requires us to locate the instantaneous event of making contact at the instant where two intervals meet. This will always be expressible as the meeting point of two atomic intervals (moments). This example shows why it is necessary to have instants as well as moments in a discrete model; but they do not have to be postulated separately, since the existence of the instants follows of necessity from the fact of each moment's immediately preceding the next.

4 Continuity

We have mentioned continuity several times; in this section we examine more closely what it entails. Two different notions of continuity are suggested by our experience of the physical world. On the one hand there is the *continuity of space and time*, which we are accustomed to regard as "seamless" continua admitting arbitrarily fine subdivision and no "gaps". On the other hand, there is the *continuity of change*, according to which measurable physical magnitudes such as the position, velocity, acceleration, or temperature of a body vary smoothly in time, again presenting an appearance of seamlessness, with no instantaneous jumps.

Notice here that we only refer to the *appearance* of seamlessness. Of course, many phenomena which appear continuous at one scale are seen to be discrete when observed more closely. An obvious example is afforded by the atomic structure of matter. Another example, with a temporal dimension, is provided by the illusion of continuity produced by the rapid succession of frames in a cine-film. As mentioned in the introduction, we are working with idealisations of the world we experience, and our purpose in this section is to examine closely the relationship between those idealisations in which phenomena are naturally represented as continuous (e.g., using the real number system) and those—such as the qualitative system of RCC-8—in which continuity does not find an obvious or natural expression.

Continuity of space and time is represented mathematically by modelling space and time in terms of the ordered set $(\mathbb{R}, <)$ of real numbers. The time dimension is represented by $\mathbb{R}$ itself, each number corresponding to a temporal instant; space is represented by the Cartesian product $\mathbb{R}^3$, each triple of real numbers corresponding to a single spatial point. The special features of $(\mathbb{R}, <)$ which suit it for this role are

- *Density*: between any two real numbers there is a third, and hence, by iteration, infinitely many. (Cf. our axiom DEN.)
- *Dedekind completeness*: if $\mathbb{R}$ is partitioned into two disjoint subsets L and R such that every member of L is less than every member of R, then either L has a greatest member, or R has a least member, but not both.

Note that neither the integers nor the rational numbers possess both these properties. The integers fail on both counts, the rationals on only the second. In the absence of Dedekind completeness, the temporal sequence admits 'gaps', at which the totality of instants can be divided into two parts L and R without there being a unique instant to mark the point of division; this is felt to be incompatible with continuity.

Temporal intervals and spatial regions, on this picture, must be specified in terms of the relationships they bear to the instants or points that have been identified with real numbers or triples thereof. From a physical point of view, we require each point (or instant) P to bear exactly one of the following relations to each region (or interval) R:

- P is *inside* R;
- P is *on the boundary of* R;
- P is *outside* R.

We shall regard an interval/region as entirely determined once it is known for every instant/point whether it lies inside, outside or on the boundary of the interval/region.

Note that this criterion of identity for intervals and regions is blind to the open/closed issue. There is no physical significance to the idea of a membership relation that is separate from the inside/outside/boundary trichotomy. This trichotomy *excludes* the idea that a region can 'contain its own boundary', there being no separate notion of containment apart from 'inside'. Nothing is gained by identifying a region either with the set of its interior points or with the set of its interior points plus boundary points, since a set is an abstract notion such that the relation it bears to its members is *sui generis* and not to be confused with the relation between a whole and its parts or between a region and the points of its interior or boundary.

The second kind of continuity, continuity of change, is modelled mathematically by representing measurable magnitudes as functions (of time) that are continuous in the special mathematical sense, namely:

> The function $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous at the point $x_0 \in \mathbb{R}$ so long as, for every real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that for every $x \in \mathbb{R}$, if $x_0 - \delta < x < x_0 + \delta$ then $f(x_0) - \epsilon < f(x) < f(x_0) + \epsilon$.

The intuitive notion of continuity demands that a function should be continuous at *every* point: so there are 'no jumps'. This excludes certain well-known 'pathological' cases, e.g., a function which is continuous at irrational points but not at rational ones.

Granted that the mathematical model is well suited to modelling continuous change, there is now a problem about modelling *discontinuous* change. This is the classical 'dividing instant' problem, which goes back to Plato and Aristotle. It may be phrased as follows: Let S be a state such that $Holds(-S, i) \wedge i|j \wedge Holds(S, j)$; can we determine whether $Holds\text{-}at(S, i][j)$?

An easy instance of this problem is the following: let S be the state $pos(a) = r$. We are supposing that an interval throughout which the position of a is different from r meets an interval throughout which a is in position r:

$$Holds(pos(a) \neq r, i) \wedge i|j \wedge Holds(pos(a) = r, j).$$

For $pos(a) \neq r$, it is enough that the space occupied by some part of a is disjoint from r—in particular, a does not have to be right outside r to count as not being at r. If the motion of a is continuous, then for a to move to r from any position r' distinct from r it must pass through a range of intermediate positions forming a path from r' to r.

Suppose, then, that a is not at r at the instant $t = i][j$. Let u be any instant dividing j: since a is at r thoughout j, it is at r at u. So a moves from r' at t to r at u. Hence, by continuity, it must occupy positions along a path joining r' to r at some times in the interval $\langle t, u \rangle$. But this is a subinterval of j. Hence there are times during j when a is not at r—contradicting our assumption that a is at r throughout j. It follows that a must be at r at instant t as well. There is an asymmetry between the states represented by 'a is at r' and 'a is not at r', which could be expressed by saying that the former must be true on closed sets of instants, the latter on open sets (compare the *continuity rule* of (Williams 1990)). In the somewhat infelicitous terminology of Galton (1990), 'a is at r' is a *state of position* whereas 'a is not at r' is a *state of motion*.

In general, if the state S describes the state of the world with respect to some continuously variable property (such as the position of a body), then the above type of argument can be used to determine a solution to the dividing instant problem. The problem becomes more vicious when S is essentially discontinuous, i.e., when there are no intermediate states between a state of the world described by S and a state of the world described by $-S$.

An example which is often cited in this connection concerns a lamp which may be on or off. Suppose the lamp is off over the interval $i = \langle t_1, t_2 \rangle$ and on over $j = \langle t_2, t_3 \rangle$. The problem is whether the lamp is on or off at the instant $t_2 = i][j$. There are a number of different responses one might make here:

1. The lamp is neither on nor off at t_2. This response comes in two flavours:
   (a) The proposition 'The lamp is on at t_2' is ill-formed—propositions have truth-values over intervals, not at instants.
   (b) The proposition does have a truth-value—but it is neither true nor false. Instead we use a three-valued logic, and assign the third truth value to this case.
2. The lamp is both on and off at t_2. This is bizarre, but I have heard it seriously suggested.
3. The lamp is off at t_2. To justify this, we treat the state of the lamp as continuously variable: we postulate a real variable l, taking values in the range $[0, L]$ and representing the amount of illumination from the lamp. Then 'The lamp is off' means $l = 0$, while 'The lamp is on' means $l > 0$ (this case is exactly parallel to that of our body a moving *away* from the position r).
4. The lamp is on at t_2. This can be similarly justified by means of a different interpretation of 'on' and 'off': this time say that 'The lamp is off' means $l < L$, and 'The lamp is on' means $l = L$. This case is parallel to a moving *to* position r, exactly as in our first example.
5. Finally, one might reject the premisses, and deny that it is possible for an interval over which the lamp is off to be immediately followed by an interval over which it is on. To justify this, one must deny that 'The lamp is off' is the negation of 'The lamp is on'. We can do this by defining 'The lamp is off' to mean $l = 0$,

and 'The lamp is on' to mean $l = L$. In order for the lamp to change from being off to being on, it has to pass over the range of values for which $0 < l < L$, and this must take time. What happens is that the lamp is off over some interval $\langle t_1, t_2 \rangle$, it is on over an interval $\langle t_3, t_4 \rangle$, and neither on nor off over the interval $\langle t_2, t_3 \rangle$, which may be of extremely short duration. Of course the dividing instant problem arises again with respect to the instants t_2 and t_3, but this time it is of the relatively harmless continuous variety that we have dealt with already: the lamp is off (i.e., $l = 0$) at t_2, and on (i.e., $l = L$) at t_3.

Response 1(a) is interesting because it has become prominent in AI, largely owing to the influence of Allen. Allen's work was prefigured by that of Hamblin in philosophy (Hamblin 1969, Hamblin 1971). Hamblin was very much motivated by the dividing instant problem. Here we highlight the fact that there is a radical incompatibility between the view of the world espoused by Allen and Hamblin and that implicit in the standard mathematical view which models continuity using the real numbers.

This incompatibility does not only arise from the dividing instant problem. A more serious problem concerns what I call *instantaneous tenure* (Galton 1994). By *tenure* of a state I mean an event which consists of that state's holding for a certain time, flanked by times at which it does not hold. For example, if the lamp is off over interval $\langle t_1, t_2 \rangle$, on over $\langle t_2, t_3 \rangle$, and off again over $\langle t_3, t_4 \rangle$, then an event of tenure of the state of the lamp's being on occurs on the interval $\langle t_2, t_3 \rangle$. Hamblin explicitly denied the possibility that a tenure event could be instantaneous, referring to the impossibility of a red book turning green just at midnight and then immediately becoming red again, so that there is only a single instant at which it is green. We may grant Hamblin his example, but what of the ball tossed up into the air: surely there is an instant, and only an instant, at which it is moving neither up nor down? On the standard mathematical view, this is inescapable. Aristotle would have said that in this situation what actually happens is that the ball comes to rest at the highest point of its trajectory, *stays there for a short interval*, then begins descending. If this is correct, then there is a serious mismatch between what actually happens and the mathematical apparatus standardly used by physicists to describe what happens in cases like this. (Note that this issue cannot, in principle, be settled by observation or measurement, since the length of the supposed interval might always be below the threshold of discrimination; once again, the matter at stake is one of finding a workable idealisation.)

Even more seriously, consider the case of a body a moving uniformly along a line from position p to position r, passing through position q on the way. This is the example used to criticise Allen in (Galton 1990). Suppose the whole movement takes up interval $\langle t_1, t_3 \rangle$. The part of the movement during which a moves from p to q must occupy some initial interval of this, say $\langle t_1, t_2 \rangle$. The remainder of the interval, namely $\langle t_2, t_3 \rangle$, is taken up by the movement from q to r. It is natural to say that a is at q at instant t_2. On the other hand, it is not at q at any other time during $\langle t_1, t_3 \rangle$. We thus have instantaneous tenure of the state of a's being at q.

We cannot dismiss this by claiming that it does not matter what we say is the case at t_2, for if we are not prepared to say that a is at q at t_2, we must deny that a is at q at all during the interval $\langle t_1, t_3 \rangle$, despite the explicit supposition that a passes through q. We should have to allow that a body can *pass through* a position without *being* there.

The alternative is to say that if a passes through q then it must spend some time there; but this destroys continuity. For there is not enough time for a to spend a positive duration at each of a nondenumerable set of positions along the path from p to r, so either we have to deny that the path contains nondenumerably many positions (which entails denying Dedekind completeness, and hence the continuity of space) or we must deny that a occupies every position along the path at some time during its move (which entails denying continuity of movement).

If we are committed to the beliefs that (a) all motion must be continuous, and (b) a body cannot occupy a position without spending some positive length of time there, then the only way out is to deny that motion is possible at all. Perhaps something of this sort was what motivated Zeno's arrow paradox, though of course Zeno was working with a conception of continuity that had not yet been formulated in terms of the Dedekind property.

5 The Theory of Dominance

When giving the occurrence conditions for $connect$, we noted that if we have DC over i and PO over j, where i meets j, then we must, by continuity, have EC at the instant $i][j$. In this section we explore further the role that continuity has to play in arguments of this kind. Essentially, we shall look at the structure of the state-space consisting of the RCC-8 relations from the point of view of its being a qualitative projection of an underlying continuous space.

We say that a state S' is a *perturbation* of state S if and only if one of these states can hold at an instant which limits an interval throughout which the other state holds, i.e., at least one of the following situations can occur:

$$Holds(S,i) \wedge Lim(t,i) \wedge Holds\text{-}at(S',t)$$
$$Holds(S',i) \wedge Lim(t,i) \wedge Holds\text{-}at(S,t)$$

If only the first of the above situations can occur, then we shall say that state S' *dominates* state S, written $S' \succ S$, whereas if only the second can occur, S' is dominated by S, written $S' \prec S$. The motivation for the term 'dominance' is as follows: suppose that S holds throughout i and that S' holds throughout j, where i meets j. Then we can think of S and S' as being in competition as to which of them, if either, should hold at the instant $i][j$. It is the dominant state which wins.

To illustrate these ideas, we divide the state-space for a single real variable into the three qualitative states positive, negative and zero, abbreviated P, N, and Z respectively. Assuming continuous variation, we have

$$Holds(Z,i) \wedge Lim(t,i) \rightarrow Holds\text{-}at(Z,t).$$

This is because if the value of the variable is non-zero at t, then by continuity it must assume values arbitrarily close to that non-zero value, and hence non-zero themselves, at all times sufficiently close to t, and this is incompatible with its being zero throughout an interval beginning or ending at t. This means that neither of the situations

$$Holds(Z,i) \wedge Lim(t,i) \wedge Holds\text{-}at(P,t)$$
$$Holds(Z,i) \wedge Lim(t,i) \wedge Holds\text{-}at(N,t)$$

can occur. On the other hand, if our variable increases uniformly over the interval $\langle -1, 1\rangle$ so that its value is 0 at time 0, then we have

$$Holds(N, \langle -1, 0\rangle) \wedge Lim(0, \langle -1, 0\rangle) \wedge Holds\text{-}at(Z, 0)$$
$$Holds(P, \langle 0, 1\rangle) \wedge Lim(0, \langle 0, 1\rangle) \wedge Holds\text{-}at(Z, 0).$$

It follows that Z dominates both N and P. Moreover, N and P are not perturbations of each other (and hence neither can dominate the other), since it is not possible for the value of the variable to change from N to P or vice versa without passing through Z. The complete dominance relation on the set $\{N, Z, P\}$ is therefore given by $Z \succ N \wedge Z \succ P$.

While it is quite possible two specify two states which are mutual perturbations although neither dominates the other, we shall only be concerned with state spaces in which all perturbation relations involve dominance in one direction or other, as in the example above. The general definition is:

A *dominance space* is a pair $(\mathcal{S}, \succ)$, where

- $\mathcal{S}$ is a finite set of *states*
- $\succ$ is an irreflexive, asymmetric relation on $\mathcal{S}$, where $S' \succ S$ is read "S' dominates S",

and the following *temporal incidence rule* holds

$$\forall S, S' \in \mathcal{S}(Holds(S, i) \wedge Lim(t, i) \wedge Holds\text{-}at(S', t) \rightarrow S' \succeq S)$$

(where $S' \succeq S$ abbreviates $(S' \succ S) \vee (S' = S)$).

The temporal incidence rule ensures that if a state S holds throughout an interval i then the only states apart from S itself which can hold at the limits of i are ones which dominate S.

The key fact about dominance spaces is that a set of such spaces can be combined into a composite dominance space, as shown by the following theorem.

Theorem 1. Let $(\mathcal{S}_1, \succ_1)$, $(\mathcal{S}_2, \succ_2)$, $\ldots$, $(\mathcal{S}_n, \succ_n)$ be dominance spaces. Then

$$(\mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n, \succ)$$

is also a dominance space, where $\succ$ is defined by the rule

$\vec{S} \succ \vec{S'}$ if and only if $S_i \succeq S'_i$ for $i = 1, \ldots n$ and $\vec{S} \neq \vec{S'}$,

where $\vec{S}$ denotes the ordered n-tuple $(S_1, S_2, \ldots, S_n)$, understood as representing the state-conjunction $S_1 \sqcap S_2 \sqcap \cdots \sqcap S_n$.

Proof. First, since $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_n$ are all finite, so is $\mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$. Next, we must check the properties of $\succ$. That $\succ$ is irreflexive follows immediately from the condition $\vec{S} \neq \vec{S'}$ appearing in the definition. To show that $\succ$ is asymmetric, suppose that both $\vec{S} \succeq \vec{S'}$ and $\vec{S'} \succeq \vec{S}$. Then for $i = 1, 2, \ldots, n$, both $S_i \succeq S'_i$ and $S'_i \succeq S_i$, so by asymmetry of $\succ_i$, $S_i = S'_i$. It follows that $\vec{S} = \vec{S'}$. Hence if we have $\vec{S} \succ \vec{S'}$ then we do *not* have $\vec{S'} \succ \vec{S}$.

Finally, we must check the temporal incidence rule. Suppose that

$$Holds(\vec{S}, i) \wedge Lim(t, i) \wedge Holds\text{-}at(\vec{S'}, t).$$

Then for $k = 1, 2, \ldots, n$ we have

$$Holds(S_k, i) \wedge Lim(t, i) \wedge Holds\text{-}at(S'_k, t),$$

which by the kth temporal incidence rule implies that $S'_k \succeq_k S_k$. We thus have

$$(S'_1 \succeq_1 S_1) \wedge (S'_2 \succeq_2 S_2) \wedge \cdots \wedge (S'_n \succeq_n S_n),$$

i.e., $\vec{S'} \succ \vec{S}$ as required. □

The importance of this theorem is that it enables us systematically to build complex dominance spaces from simpler ones. We can start with very simple spaces where the dominance relations are easy to verify 'by hand', and then build up to more complicated cases where these relations are less straightforward to determine. Since we are dealing with dominance spaces, information regarding dominance also provides complete information about perturbation as well. In the next section we illustrate this with two examples of particular relevance to the present study. Further examples can be found in (Galton 1995).

6 Applications to Spatial Change

We can construct RCC-8 as a dominance space by noting that the RCC-8 relations are uniquely determined by knowing (a) whether all, some, or none of A is inside B, (b) whether all, some, or none of B is inside A, and (c) whether A and B share any boundary points. (Here we are using the word 'some' in its exclusive sense, i.e., some but not all). This representation is related, but not identical, to Egenhofer's 4-intersection. The three factors can be represented numerically as follows:

(a) Let

$$\alpha = \frac{\text{area of A inside B}}{\text{area of A}}.$$

Then the three states are $\alpha = 0$ (none of A is inside B), $0 < \alpha < 1$ (some of A is inside B), and $\alpha = 1$ (all of A is inside B). The state 'some' is dominated by both 'none' and 'all'.

(b) As with (a), but with A and B swapped around:

$$\beta = \frac{\text{area of B inside A}}{\text{area of B}}.$$

(c) Let γ be the minimum distance between a boundary point of A and a boundary point of B. Then we have two states $\gamma = 0$ (i.e., A and B share at least one boundary point) and $\gamma > 0$ (A and B do not share any boundary point). The state $\gamma = 0$ dominates the state $\gamma > 0$.

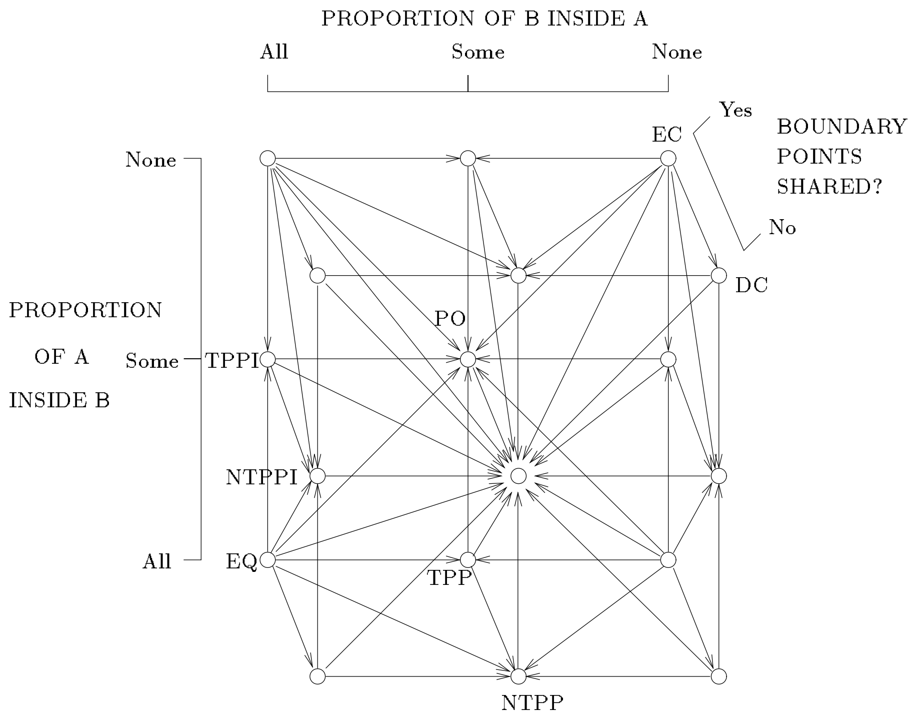

Figure 1: *Generation of the dominance space for the RCC-8 spatial relations.*

These three little dominance spaces would, if they were mutually independent, combine to give a space with $3 \times 3 \times 2 = 18$ elements. The dominance relations are computed using Theorem 1, as shown in Figure 1, where the arrows indicate the direction of dominance.

In fact the three spaces are not independent; all but 8 of the elements are impossible (e.g., if none of A is inside B then none of B can be inside A either). The eight possible combinations correspond exactly with the spatial relations in RCC-8, as follows[1]:

	B in A	A in B	Share boundary
DC	none	none	no
EC	none	none	yes
PO	some	some	yes
TPP	some	all	yes
NTPP	some	all	no
TPPI	all	some	yes
NTPPI	all	some	no
EQ	all	all	yes

[1]Note that this analysis assumes that each region consists of a single connected component.

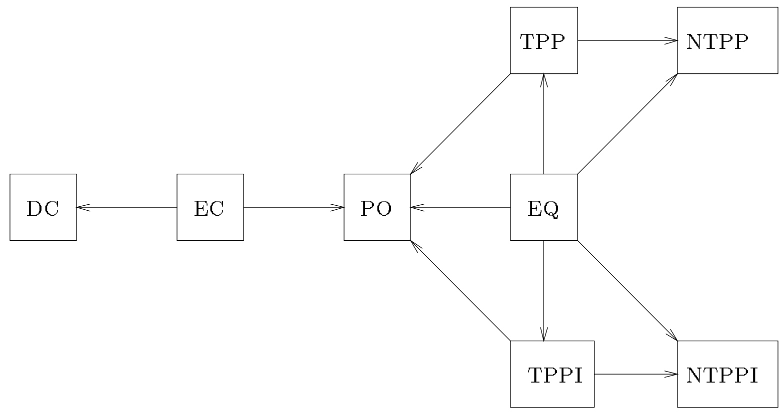

Figure 2: *The dominance space for the RCC-8 spatial relations.*

In Figure 1 the nodes corresponding to these eight possibilities are labelled with the appropriate RCC-8 designations. A clearer view of the dominance space can be obtained by deleting all the 'impossible' nodes and rearranging those that remain as in Figure 2. It is reassuring to note that this diagram is identical, apart from the addition of the dominance arrows, to that of (Randell et al. 1992)!

From this figure we can read off, for example, that a transition between DC and PO must involve passing through EC, though since this dominates both its neighbours the intermediate state need only hold for an instant. We can also see that the state TPP can hold for an single instant in the context of a transition between PO and NTPP (since it dominates both of these states), but if it holds in the context of a transition between EQ and NTPP then it must do so for an interval (since it is dominated by EQ). These conclusions are in conformity with the demands of continuity.

For a more complicated example, we consider the position of a non-rigid body in relation to two fixed disconnected regions. There are eight relations which the position of the body can stand in with respect to either region individually. If its position with respect to one region were free to vary independently of its position with respect to the other, this would give us a state space containing 64 product relations. In fact only 31 elements of the full Cartesian product can be realised. In particular, if the position of the body in relation to one region is EQ, TPP, or NTPP, then it can only be DC with respect to the other. This is because the two regions are themselves DC. This gives us the six composite relations (EQ,DC), (TPP,DC), (NTPP,DC), (DC,EQ), (DC,TPP), and (DC,NTPP). The remaining five simple relations are genuinely independent for each of the two positions, giving a further 25 composite relations. All 31 relations are portrayed pictorially in Figure 3, with the dominance—and hence perturbation—relations as determined from Theorem 1.

Using this figure, we can observe that, for example, if a body is EC to region 1 and DC to region 2, then in order to become DE to 1 and EC to 2 it must first become either

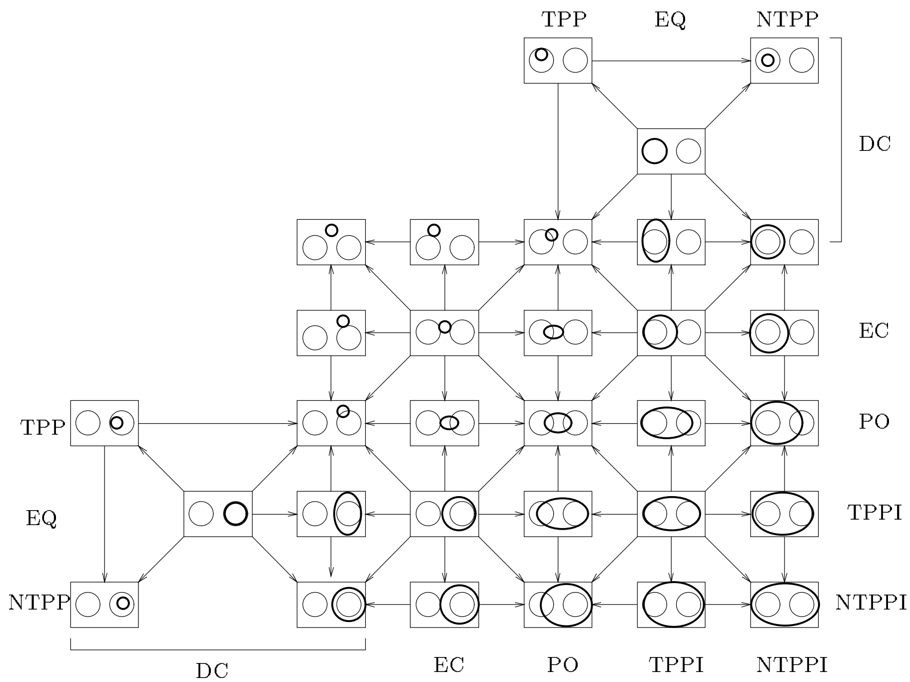

Figure 3: *Position of a non-rigid body in relation to two fixed regions.*

DC to both or EC to both—the latter case being possible for a single instant, the former requiring an interval.

7 Concluding Remarks

We have provided a theoretical basis for describing movement events in terms of the conditions for their occurrence, which consist of the holding or not holding of various positional fluents at different times. To do this we have brought together a formal model of time, based on a set of instants endowed with a total ordering, with a formal model of space based on regions. By attending closely to the constraints imposed by continuity on the temporal behaviour of different fluents we developed a theory of *dominance* by which we are able to generate from first principles the neighbourhood relation on the RCC-8 set of qualitative positional relations, as well as on more complicated state-spaces such as the possible positions of a non-rigid body in relation to two fixed regions.

We advocate the use of dominance as providing a systematic tool for investigating the structure of qualitative state-spaces derived from an underlying space that is conceptualised as continuous. As such, the theory of dominance is not only of interest in the spatial domain, although that does provide a highly appropriate testing ground, with perhaps the richest set of particular examples, for the theory. There is considerable scope for further research in investigation of the dominance structure of other qualitative spaces in the literature on spatial reasoning, as for example the orientation-based system of (Freksa 1992).

With regard to the work presented in this paper, an intriguing open question remains. Our theory of dominance depends on our being able to speak meaningfully of a state holding at an instant. We used the theory to help us generate the perturbation and dominance relations for complex state-spaces. The resulting diagrams can be read as supplying information about perturbation only, ignoring the dominance information. As such, they can be recognised as valid *even if one does not accept the premiss that states can be said to hold at instants*. The utility of dominance in deriving results relating to perturbation seems to indicate that dominance expresses deep-seated characteristics of the spaces in question. But as explicitly presented here, it is incompatible with the widespread view (as expressed for example by Allen and Hayes (1985)) that it does not make sense to speak of a state holding at an instant, but only over an interval. The question facing us is therefore whether there exists an alternative formulation of the dominance theory which can play the same role as that theory in uncovering the perturbation structure of complex state-spaces, without presupposing the contentious association of states with instants.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals, *Communications of the ACM* **26**: 832–843.

Allen, J. F. (1984). Towards a general theory of action and time, *Artificial Intelligence* **23**: 123–154.

Allen, J. F. and Hayes, P. J. (1985). A common-sense theory of time, *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pp. 528–531.

Egenhofer, M. J. (1991). Reasoning about binary topological relations, *in* O. Günther and H.-J. Schek (eds), *Advances in Spatial Databases*, Springer-Verlag, pp. 143–160.

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning, *in* A. U. Frank, I. Campanari and U. Formentini (eds), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Springer, Berlin, pp. 162–178. Lecture Notes in Computer Science, Volume 639.

Galton, A. P. (1990). A critical examination of Allen's theory of action and time, *Artificial Intelligence* **42**: 159–188.

Galton, A. P. (1993). Towards an integrated logic of space, time, and motion, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*.

Galton, A. P. (1994). Instantaneous events, *in* Hans Jürgen Ohlbach (ed.), *Temporal Logic: Proceedings of the ICTL Workshop*, Max-Planck-Institut für Informatik, Saarbrücken.

Galton, A. P. (1995). A qualitative approach to continuity, *Time, Space, and Movement: Meaning and Knowledge in the Sensible World.* Working papers of TSM'95, Château de Bonas, 23–27 June 1995, Groupe 'Langue, Raisonnement, Calcul', Toulouse.

Hamblin, C. L. (1969). Starting and stopping, *The Monist* **53**: 410–425.

Hamblin, C. L. (1971). Instants and intervals, *Studium Generale* **24**: 127–134.

Randell, D. A., Cui, Z. and Cohn, A. G. (1992). A spatial logic based on regions and connection, *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, pp. 165–176. Cambridge, Massachusetts, October 1992.

van Benthem, J. F. A. K. (1983). *The Logic of Time*, Kluwer, Dordrecht. Second Edition, 1991.

Williams, B. C. (1990). Temporal qualitative analysis: Explaining how physical systems work, *in* D. S. Weld and J. de Kleer (eds), *Qualitative Reasoning about Physical Systems*, Morgan Kaufmann, San Mateo, California, pp. 133–177.