Carving Up Space: steps towards construction of an absolutely complete theory of spatial regions*

Brandon Bennett**

Division of Artificial Intelligence
School of Computer Studies
University of Leeds, Leeds LS2 9JT, England
brandon@scs.leeds.ac.uk

Abstract. Motivation is given for the construction of an absolutely complete theory of spatial regions. Additional axioms for the RCC theory (Randell, Cui and Cohn 1992) are suggested to restrict the class of models satisfying this theory. Specific problems addressed are the characterisation of dimension and the provision of adequate existential axioms.

1 Introduction

Spatial information is essential to a broad spectrum of knowledge domains and to many important reasoning tasks. Hence, an ontology of space and spatial regions is fundamental to the development of representations and reasoning mechanisms for AI systems.

Geometry and Topology are both highly developed fields of mathematics. In both areas, the formal theories developed take *points* (or in the case of incidence geometry, points and lines) as primitive elements from which objects corresponding to regions are constructed set-theoretically. From the point of view of knowledge representation and automated reasoning this often leads to difficulties. One problem is that the most natural and useful way of presenting many kinds of spatial information is in terms of relationships that hold between *regions* of space or the bodies that occupy those regions. Another is that the use of set theory leads to highly intractable formal systems.

Although region based formalisms have received relatively little attention, a number of significant theories have been produced (de Laguna 1922, Whitehead 1929, Leonard and Goodman 1940, Tarski 1956). The need for detailed formal analysis of spatial information has been recognised by researchers in AI. Clarke's theory (Clarke 1981, Clarke 1985) has been taken up and modified by Randell, Cohn and Cui (Randell and Cohn 1989, Randell et al. 1992) and also, more recently, by Asher and Vieu (1995).

* This work was supported by the EPSRC under grant GR/K65041.

** Contributions to this paper by Dr A.G. Cohn and Dr N.M. Gotts and suggestions from O. Lemon are gratefully acknowledged.

Decidability and complexity issues have so far been addressed only for small sub-domains of region-based spatial reasoning. Bennett (1994) gives a decision procedure for testing consistency of sets of binary topological relations drawn from a limited but fairly expressive set (including all the RCC-8 relations shown in figure 1 of section 1.3). The complexity of this decision problem was later shown by Nebel (1995) to be polynomial in the number of relations being tested.[3]

1.1 Absolutely Complete Theories

The purpose of this paper is to motivate the construction of a *complete* theory of spatial regions and present some significant steps towards its development. A logical theory is (absolutely) complete if for every formula ϕ expressed in the vocabulary of the theory (i.e. some fixed collection of relation and function symbols together with the usual logical operators) either ϕ or $\neg\phi$ is a consequence of the axioms of the theory.

Contemporary use of the term 'completeness' often refers to a weaker property of theories which depends upon having a model-theoretic semantics in which the theory is interpreted. A theory is said to be complete, relative to the semantics, if all formulae that are true in every model are provable from its axioms. Under this relative conception of completeness, there will generally be *contingent* formulae, which are true in some models and false in others and if this is the case then the theory will not be complete in the absolute sense. In an absolutely complete theory, every formula is either true in every model or false in every model.

A related property is *categoricity*: a theory is *categorical* if all its models are isomorphic. Clearly if this is the case then the theory must be complete; but it is possible that a theory have non-isomorphic models and yet exactly the same formulae are true in every model — so it is complete but not categorical. In fact, if a 1st-order theory has denumerable models there will always be larger (unintended) models because denumerability is not 1st-order definable. Hence, the concept of $\aleph_0$-*categoricity* is often more useful. This means that all countable models are isomorphic.

Complete theories are desirable in knowledge representation for two main reasons. Firstly, from the point of view of the semantic adequacy of a theory, a complete theory may be regarded as preferable to an incomplete one in that it may be regarded as fixing the meanings of all non-logical constants. Secondly, from the point of view of automated reasoning, every complete theory has the important property of being *decidable* — i.e. there is a procedure which will determine in finite time whether any formula is valid. This follows from the semi-decidability of any finitely (or recursively) axiomatisable theory: any valid formula is provable in a finite number of steps, so, because in a complete theory every formula is either valid or its negation is valid, by concurrently trying to prove both ϕ and $\neg\phi$ we have a decision procedure. (Of course, a decision procedure constructed in this way will almost certainly be totally impractical; but, if a useful theory is

[3] More precisely, the problem is in the complexity class NC, which can be computed in polylogarithmic time by the use of polynomially many processors.

known to be decidable on the grounds of its completeness, it will be worthwhile seeking a more computationally viable decision procedure.)

For many purposes (e.g. representing some real world situation) it is convenient to augment the vocabulary of a theory with additional constant terms denoting specific objects. Clearly, in this extended language, formulae containing such constants may be *contingent*, even though all relational and functional vocabulary is constrained by an absolutely complete theory. (This may be one reason why the importance of absolute completeness is often overlooked by those interested in knowledge representation.) However, even when we are interested in this wider class of formulae, absolute completeness is a valuable property. Suppose we have an absolutely complete theory and a formula containing, in addition to the vocabulary of the theory, certain non-theoretical constants: by replacing such constants with variables which are existentially quantified (with widest scope) we get a formula that is a provable consequence of the theory iff the original formula is *possible* with respect to the theory (and provably false if it is impossible); and, if alternatively we replace the constants with universally quantified variables, we get a formula that is provable iff the original formula is a *necessary* consequence of the theory (and provably false otherwise). Hence, relative to an absolutely complete theory, in reasoning about formulae which may contain additional non-theoretical constants, there is a decision procedure which will distinguish between necessary, contingent and impossible formulae.

1.2 Deficiency of Region-Based Theories

Complete axiom systems are known for several point-based geometries (see e.g. (Tarski 1959)). The models of such systems are isomorphic to fields over Cartesian tuples of numbers which may be regarded as the coordinates of points. But for region based spatial theories this kind of analysis is not available and there is a lack of meta-mathematical and model-theoretic results about proposed formalisms. The systems have for the most part been presented as uninterpreted calculi, with models being suggested only to give some intuitive understanding of the primitive concepts.[4] This problem has only recently been addressed: Biacino and Gerla (1991) have shown the relationship between the theories of Leonard and Goodman and of Clarke to the well-known mathematical structures of Boolean algebra and (ortho-complemented) lattice; and Asher and Vieu (1995) have presented a mereo-topological theory which is shown to be sound and complete with respect to a certain class of model structures.

(Tarski 1956) is the only known theory of spatial regions that is *complete* and ($\aleph_0$-)*categorical*. Tarski's theory is only made categorical by indirect means: firstly the notions of *point*, *equidistance* and *betweenness* are introduced by a series of definitions; then it is stipulated that these defined concepts obey the axioms of Euclidean geometry (Tarski 1959). He admits that the resulting system is not ideal:

[4] Model theoretic properties of calculi of temporal intervals are much better understood than those of spatial formalisms. Allen's (1981) interval calculus has been thoroughly investigated by Ladkin (1987) and others.

> The postulate system given above is far from simple and elegant; it seems very likely that this postulate system can be essentially simplified by using intrinsic properties of the geometry of solids. (Tarski 1956)

He then gives an example of how an axiom stated indirectly in terms of points can be replaced by a simple existential axiom concerning the primitive notion of *sphere*.

A particular weakness in all the theories is the lack of attention to existential axioms. Universal properties of primitives such as 'connectedness' or the 'part/whole' relation seem to be more obvious than statements guaranteeing the existence of regions exhibiting specific properties and configurations. Existential axioms require us to make choices about what counts as a region and to be definite about the domain of regions and its structure. Accordingly they are essential in rendering a theory categorical and thus fixing a single model modulo isomorphism (and denumerability if the theory is 1st-order).

1.3 The RCC Theory

In this paper I shall take as my starting point the 1st-order[5] theory presented by Randell et al. (1992) — henceforth the 'RCC' theory. This theory is a modification of the theory of Clarke (1981, 1985) (which is in turn derived from the theories of Whitehead (1929) and Leonard and Goodman (1940)).

The basic RCC theory assumes just one primitive dyadic relation: $\mathsf{C}(x, y)$ read as 'x is connected to y' and the domain is intended to be that of spatial regions. The C relation is reflexive and symmetric, which is ensured by the following two axioms:

$$\forall x \mathsf{C}(x, x) \qquad \textbf{(Cref)}$$

$$\forall xy[\mathsf{C}(x, y) \rightarrow \mathsf{C}(y, x)] \qquad \textbf{(Csym)}$$

Many other useful relations can be defined in terms of C, for example:

- Part, $\mathsf{P}(x, y) \equiv_{\text{def}} \forall z[\mathsf{C}(z, x) \rightarrow \mathsf{C}(z, y)]$
- Proper Part, $\mathsf{PP}(x, y) \equiv_{\text{def}} \mathsf{P}(x, y) \wedge \neg\mathsf{P}(y, x)$
- Overlap, $\mathsf{O}(x, y) \equiv_{\text{def}} \exists z[\mathsf{P}(z, x) \wedge \mathsf{P}(z, y)]$
- External Connection, $\mathsf{EC}(x, y) \equiv_{\text{def}} (\mathsf{C}(x, y) \wedge \neg\mathsf{O}(x, y))$
- Non-Tangential Proper Part, $\mathsf{NTPP}(x, y) \equiv_{\text{def}} \mathsf{PP}(x, y) \wedge \forall z[\mathsf{C}(z, x) \rightarrow \mathsf{O}(z, y)]$

Figure 1 shows eight definable relations which constitute a jointly exhaustive and pairwise disjoint (JEPD) set — i.e. for any pair of regions exactly one of these relations holds.

The theory also contains some 'quasi-Boolean' functions characterised by further axioms. These are discussed below in section 2.2. It is in the quasi-Boolean function axioms that the most significant differences between the RCC theory and the theory of Clarke lie.

[5] In fact, as we shall see in section 2.2, a *sorted* 1st-order logic is employed; but we don't need to worry about that yet.

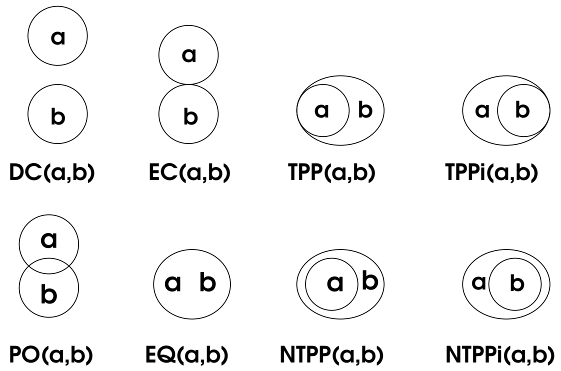

Fig. 1. RCC-8 — a set of JEPD topological relations

2 Existence

A complete theory of spatial regions must not only provide a means of stating and reasoning about relationships holding between spatial regions. It must also provide an ontology capable of determining the existence or otherwise of any configuration of regions describable in the vocabulary of the theory.

Some definitions of relations in terms of C involve existential quantifiers. This quantification guarantees the existence of certain regions in order to *witness* the fact that some relation holds between other regions. E.g., if we know that two regions overlap, there must be some region which is part of both regions. However, such existential commitment is contingent on certain kinds of fact being asserted. No existential facts follow *a priori* from the symmetry and reflexivity of C.

The existential component of a complete theory of space must determine exactly what regions exist. In the next few sections I investigate what regions we might expect to exist and present axioms to ensure their existence.

2.1 Identity and Extensionality

Before considering questions of existence of regions we need a clear idea of their identity conditions. Axiomatic theories (particularly those which seek to characterise a single primitive relation), often contain some kind of *axiom of extensionality.* This is an axiom which asserts that identity of any two objects follows from their indiscernibility with respect to some property. In the RCC calculus of regions the obvious axiom of extensionality would be:

$$\forall x \forall y[\forall z[\mathsf{C}(x,z) \leftrightarrow \mathsf{C}(y,z)] \rightarrow (x = y)] \qquad \textbf{(Cext)}$$

Although this axiom is given in (Clarke 1981) it is not part of the theory given in (Randell et al. 1992). It turns out that this formula is actually a consequence of another RCC axiom asserting that any two regions have a unique 'sum' (the

QBfun1 axiom given in section 2.2 below); however, in view of its fundamental importance it might be better to regard it as a basic axiom of the theory.[6].

2.2 Demarcation and Existence

The domain of geometry can be characterised by reference to the construction of geometrical figures. More specifically, $2D$ Euclidean geometry can be regarded as the theory of configurations of points and lines constructible on a plane with the aid of a ruler and compass (see (Tarski 1959)). The axioms of elementary geometry have been arrived at by considering basic operations in the construction of figures. E.g., given any two distinct points one can introduce a new point which lies between (using the ruler) or is equidistant from (using the compass) the initial two points. In this section I indicate how a similar analysis of the way in which configurations of regions may be drawn can lead us to axioms for the existence of spatial regions. My examination will be confined to $2D$ regions but, with a little more imagination and considerably more technical complication, this approach could equally well be applied to $3D$ space.

If a number of regions are demarcated by drawing their boundaries on a piece of paper, there is a sense in which the number of regions created is greater than the number of regions explicitly drawn. Other areas of the paper are demarcated indirectly. E.g., in outlining any region one automatically creates its complement, a region consisting of all of space (or all of the paper) except that which is outlined. And, for any two regions demarcated one can consider their combined space as another region.

In constructing a complete theory of spatial regions we are not concerned with any particular diagram showing a configuration of regions but rather with all possible such constructions. Thus regions exist whether or not we have actually demarcated them on a piece of paper. Nevertheless, it seems clear that the existence of a particular finite configuration of regions must mean that it would be possible to construct a figure demarcating those regions. Thus the possibilities for constructing figures correspond to existential axioms concerning regions.

RCC contains a set of definitions of *quasi-Boolean* functions from regions to regions. These can be seen as generating new regions from old, in accordance with the idea that boundedness in some figure ensures existence:

QBfun1) $\forall x \forall y \forall z[\mathsf{sum}(x,y) = z \leftrightarrow \forall w[\mathsf{C}(z,w) \leftrightarrow [\mathsf{C}(w,x) \vee \mathsf{C}(w,y)]]]$

QBfun2) $\forall x \forall \mathbf{y}[\mathsf{compl}(x) = \mathbf{y} \leftrightarrow \forall z[(\mathsf{C}(z,\mathbf{y}) \leftrightarrow \neg\mathsf{NTPP}(z,x)) \wedge (\mathsf{O}(z,\mathbf{y}) \leftrightarrow \neg\mathsf{P}(z,x))]]$

QBfun3) $\forall x \forall y \forall \mathbf{z}[\mathsf{prod}(x,y) = \mathbf{z} \leftrightarrow \forall w[\mathsf{C}(w,\mathbf{z}) \leftrightarrow \exists v[\mathsf{P}(v,x) \wedge \mathsf{P}(v,y) \wedge \mathsf{C}(w,v)]]]$

QBfun4) $\forall x \forall y[\mathsf{diff}(x,y) = \mathsf{prod}(x, \mathsf{compl}(y))]$

[6] Randell et al. (1992) define the relation $\mathsf{EQ}(x,y)$ as equivalent to $\mathsf{P}(x,y) \wedge \mathsf{P}(y,x)$. If this definition is regarded as an axiom rather than a definition and if EQ is treated as logical equality then this axiom is equivalent to **Cext**

These formulae are not purely definitional because functions in 1st-order logic carry existential import: every term must denote some individual in the domain. Thus, e.g., **QBfun1** guarantees that for every two regions there exists a unique region (their sum) which is connected to all and only those regions connected to either of the original two regions. This means that all the functions apart from sum give rise to a problem because they are not total. E.g., if two regions are disjoint then there is no (non-empty) region which is their prod. One way round this is to introduce a *null* region as a possible value of the functions. However, care must be taken since in the definitions of topological relations it was assumed that quantifiers ranged over only non-empty regions (e.g. two regions overlap if they have a mutual *non-empty* part).

In (Randell et al. 1992) a *sorted logic* (Cohn 1987) is used to make the functions total. Space does not permit a full presentation of a sorted theory but for present purposes a much simplified version will suffice. We stipulate that the quantified variables appearing in bold font in **QBfun2–3** range over non-empty regions and also the null region, whereas all other variables range only over non-empty regions. So, the range of the quasi-Boolean functions includes the null-region even though this is excluded from the domain of quantification in the C axioms and definitions.

It is worth noting that although the (total) binary sum function introduced by **QBfun1** guarantees the existence of a sum of any finite set of regions, it does not guarantee the existence of all infinite sums. By contrast the theories of Tarski (1956) and Clarke (1981) contain an axiom ensuring that there is a sum of every (finite or infinite) set of regions. It can be shown that such an infinite summation principle is inconsistent with the RCC theory. By considering, for instance, the (infinite) sum of all non-tangential parts of a region, is easy to prove (making use in particular of the complement axiom, **QBfun2**) that such a region has contradictory properties.

3 Dimension

An important parameter affecting what configurations of regions are possible is the dimensionality of the space. E.g., in Euclidean $3D$ space an infinite number of non-overlapping 3D regions may meet along the length of a line (think of the segments of an orange), whereas in (Euclidean) $2D$ only 2 non-overlapping $2D$ regions can meet in this way (though an infinite number can meet at a single point). Thus if a theory is to be complete it must fix the dimensionality of the space.

3.1 Classical Definitions of Dimension

In point-based Euclidean geometry it is possible to start with a theory which has models of all dimensionalities and then fix the the dimension of the space by means of 'upper' and 'lower' dimension axioms (Scott 1959, Tarski 1959). The former describes a situation which is impossible in any lower dimension. The latter is the negation of a description of a situation which can only hold in a higher dimension. More specifically, in a theory in which an equidistance relation is definable, a space can be characterised as being of dimension n by a (1st-order) formula stipulating that that there are $n+1$ mutually equidistant points and no more than that.

Topologically a space is said to be of dimension $< n$, if every open cover $\{O_1, \ldots, O_k\}$ can be refined to a closed cover $\{C_1, \ldots, C_k\}$, such that every point in U occurs in at most $n+1$ of the C_is (Kuratowski 1972). For example, if a 2D region is covered by overlapping open patches then the region can always be covered by non-overlapping closed regions, such that each is a part of one of the regions in the original open cover and no point is shared by more than three regions. Because it involves quantification over 'open covers', which are sets of regions, this characterisation of dimension is 2nd-order with respect to any logical theory whose basic entities are regions. Moreover, the notions of open and closed regions are not definable in the RCC theory.

3.2 A 1st-order Region-Based Characterisation

An obvious way to strengthen the RCC theory so as to fix the dimensionality of its models is to define, in terms of connection, a 1st-order predicate which characterises a region as being of a certain dimension. By universally quantifying this predicate we obtain an axiom which constrains all regions in the domain to be of the same fixed dimension.[7] Inspired by the topological definition given in the last section, Gotts (1994a) takes the following approach to constructing such a predicate.

First the relation of 'layered' partial overlap is defined by

$$\mathsf{LPO}(x,y) \equiv_{\mathrm{def}} (\mathsf{PO}(x,y) \wedge \mathsf{DC}(\mathsf{diff}(x,y), \mathsf{diff}(y,x)))$$

where $\mathsf{diff}(x,y)$ is defined as equivalent to $\mathsf{prod}(x, \mathsf{compl}(y))$. Gotts then defines the notion of an *LPO cover* of a region r: this is a finite set, S, of regions such that r is part of the sum of the regions in S and any two (distinct) regions in S are related either by LPO or by DC. It is then conjectured that, for a region of dimension n, any LPO cover made up of $n+2$ regions can be refined to an LPO cover such that no region is a common part of all regions in the refinement — i.e. the common intersection of all members of the cover is null.

If Gotts's conjecture is correct, a $1D$ region can be characterised as follows (this is not the simplest characterisation). For convenience we first define

$$\mathsf{LPODC}(x,y) \equiv_{\mathrm{def}} (\mathsf{LPO}(x,y) \vee \mathsf{DC}(x,y))$$

and

$$\begin{aligned}\mathsf{LPO3}(r,x,y,z) \equiv_{\mathrm{def}} (&r = \mathsf{sum}(\mathsf{sum}(x,y),z) \wedge \\ &\mathsf{LPODC}(x,y) \wedge \mathsf{LPODC}(y,z) \wedge \mathsf{LPODC}(z,x))\end{aligned}$$

[7] Attempting to characterise a universe in which there are regions of different dimension leads to serious difficulties, consideration of which is beyond the scope of the present work.

We can then define

$$\begin{aligned}\mathsf{DIM1}(r) \;\equiv_{\text{def}}\; \forall x \forall y \forall z[\mathsf{LPO3}(r,x,y,z) \rightarrow \exists x' \exists y' \exists z'[\; & \mathsf{LPO3}(r,x',y',z') \\ & \wedge \mathsf{P}(x',x) \wedge \mathsf{P}(y',y) \wedge \mathsf{P}(z',z) \\ & \wedge \neg\exists w[\mathsf{P}(w,x') \wedge \mathsf{P}(w,y') \wedge \mathsf{P}(w,z')]]]\end{aligned}$$

To characterise a $2D$ region, $\mathsf{LPO4}(w,x,y,z)$ is defined analogously to $\mathsf{LPO3}$. The definition of $\mathsf{DIM2}(r)$ then takes a similar form to that of $\mathsf{DIM1}$(r) but using $\mathsf{LPO4}$ instead of $\mathsf{LPO3}$ and with an additional conjunct, $\neg\mathsf{DIM1}(r)$. Similar definitions can be used to characterise any dimension.

Whether these axioms do in fact provide an adequate characterisation of dimensionality is the subject of ongoing research. The conjecture is supported by consideration of a wide variety of diagrams. The most obvious route to a rigorous proof is to correlate the region-based theory with classical geometrical or point-set topological theories within which the characterisation of dimension is well-understood; but the form that such a correlation might take is still unclear. However, if we specifically want to characterise only $2D$ regions, the axioms given in the next section may be used as an alternative.

4 Planarity

If we are to have a theory of regions with a unique model then as well as fixing the dimensionality of regions we shall also have to fix the global topology of the space. In the remainder of the paper we shall be concerned with $2D$ space. Such a space could have the topology of the plane, the sphere, the torus or any other surface but in a complete theory only one of these models must be allowed. In this section I shall give axioms intended to distinguish among certain of these possibilities and in particular I shall be concerned with characterising a planar space.

We proceed by introducing some preliminary definitions. A one piece 'one-piece' (often called 'self-connected') region is defined by

$$\mathsf{OP}(a) \;\equiv_{\text{def}}\; \forall x \forall y[((\mathsf{sum}(x,y) = a) \rightarrow \mathsf{C}(x,y))]$$

(as usual the quantified variables range only over non-null regions). We then define the predicate ICON, 'interior connected', meaning that all parts of a region are connected *via* interior points:[8]

$$\mathsf{ICON}(x) \;\equiv_{\text{def}}\; \forall y[\mathsf{NTPP}(y,x) \rightarrow \exists z[\mathsf{P}(y,z) \wedge \mathsf{NTPP}(z,x) \wedge \mathsf{OP}(z)]]$$

Finally we define the related notion of 'firm external connection' holding when two EC regions share a boundary line segment:

$$\mathsf{FEC}(x,y) \;\equiv_{\text{def}}\; (\mathsf{EC}(x,y) \wedge \exists x' \exists y'[\mathsf{P}(x',x) \wedge \mathsf{P}(y',y) \wedge \mathsf{ICON}(\mathsf{sum}(x',y'))])$$

[8] This definition is a modification that given in (Gotts 1994b) and (Gotts 1994a).

We can then rule out a large class of models by asserting that the maximum number of ICON regions which are mutually FEC is four:

$$\neg \exists x_1 \exists x_2 \exists x_3 \exists x_4 \exists x_5 [\bigwedge_{1 \leq i \leq 5,\ 1 \leq j \leq 5,\ i \neq j} \mathsf{FEC}(x_i, x_j)] \qquad (\mathbf{max4FEC})$$

The constraint imposed by **max4FEC** is illustrated by figure 2, which also shows how, if the regions are OP but not all ICON, more than four may be mutually FEC (shaded areas of the same colour are parts of the same non-ICON region).

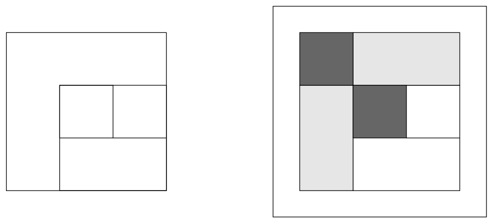

Fig. 2. Mutually FEC, OP regions in $2D$

max4FEC restricts the space to be of dimension 2 or less and also prohibits surfaces with complex topologies such as the torus (upon which a configuration of 5 mutually FEC regions can easily be constructed). **max4FEC** is however satisfied by both planar and spherical spaces.[9] So a categorical theory must contain a further axiom to select only one of these possibilities.

To make this distinction we define a predicate WCON(x) (x is 'well-connected') characterising topologically simple regions which are not only ICON but also do not contain holes. In $2D$, WCON regions can be defined as those such that they and their complement are both ICON:

$$\mathsf{WCON}(x) \ \equiv_{\text{def}} \ (\mathsf{ICON}(x) \wedge \mathsf{ICON}(\mathsf{compl}(x)))$$

We now imagine the entire space, us[10], divided into three WCON and mutually FEC parts, such that each shares a single boundary segment with the other two parts. In a planar space, these regions will meet at a single point, whereas on a sphere they will meet at two separate points. To formalise these conditions further definitions will again be useful:

We say that a WCON region r 'neatly straddles' two WCON, FEC regions, a and b, if r overlaps both a and b and is part of the sum of a and b and furthermore, the intersections of r with both a and b are also WCON:

[9] In fact **max4FEC** is also satisfied in many bounded spaces such as a disc or the surface of a cylinder. Characterisation of bounded and unbounded spaces is clearly essential to the construction of a complete theory but it is not covered in the present paper.

[10] The constant us is characterised by the RCC axiom $\forall x[\mathsf{C}(x, \mathsf{us})]$ — see section 5.4.

$$\mathsf{NS}(r,a,b) \equiv_{\text{def}} \mathsf{WCON}(r) \wedge \mathsf{WCON}(a) \wedge \mathsf{WCON}(b)$$
$$\wedge\ \mathsf{O}(r,a) \wedge \mathsf{O}(r,b) \wedge \mathsf{P}(r, \mathsf{sum}(a,b))$$
$$\wedge\ \mathsf{WCON}(\mathsf{prod}(r,a)) \wedge \mathsf{WCON}(\mathsf{prod}(r,b))$$

We can then say that regions, a and b, share a single boundary if and only if any two regions r_1 and r_2, which neatly straddle a and b, are parts of a third region r_3 which also neatly straddles a and b. I denote this relationship by $\mathsf{FEC1}(a,b)$:

$$\mathsf{FEC1}(a,b) \equiv_{\text{def}} \forall r_1 \forall r_2 [\mathsf{NS}(r_1,a,b) \wedge \mathsf{NS}(r_2,a,b) \rightarrow$$
$$\exists r_3 [\mathsf{P}(r_1,R_3) \wedge \mathsf{P}(r_2,R_3) \wedge \mathsf{NS}(r_3,a,b)]]$$

If three regions, x, y and z are mutually FEC1 and meet at a point we can stipulate that a WCON region m contains that point by saying that the intersections, $\mathsf{prod}(x,m)$, $\mathsf{prod}(y,m)$ and $\mathsf{prod}(z,m)$ are also mutually FEC1. Thus if the regions a, b and c meet at two separate points then there must be two separate (i.e. DC) regions, m_1 and m_2, both satisfying this condition. This enables us to distinguish planar or disc-like spaces from spherical spaces. We first define the mutual FEC1 relation between three regions by

$$\mathsf{MFEC1}(x,y,z) \equiv_{\text{def}} \mathsf{FEC1}(x,y) \wedge \mathsf{FEC1}(y,z) \wedge \mathsf{FEC1}(x,z))$$

Planar or disc-like universes (whose 2-dimensionality must also be fixed by an axiom such as **max4FEC**) can now be characterised by the following axiom:

$$\forall x \forall y \forall z [(\mathsf{sum}(x, \mathsf{sum}(y,z)) = \mathsf{us}) \wedge \mathsf{MFEC1}(x,y,z)) \rightarrow$$
$$\neg \exists m_1 \exists m_2 [\mathsf{WCON}(m_2) \wedge \mathsf{WCON}(m_2) \wedge \mathsf{DC}(m_1,m_2)$$
$$\wedge\ \mathsf{MFEC1}(\mathsf{prod}(x,m_1), \mathsf{prod}(y,m_1), \mathsf{prod}(z,m_1))$$
$$\wedge\ \mathsf{MFEC1}(\mathsf{prod}(x,m_2), \mathsf{prod}(y,m_2), \mathsf{prod}(z,m_2))]] \qquad \textbf{(plan)}$$

5 Carving Up Space

The quasi-Boolean functions do not tell us how to actually construct figures; they merely ensure that given a figure demarcating a number of regions, certain other (derived) regions also exist. But RCC also contains an axiom guaranteeing that every region has a non-tangential proper part:

$$\forall x \exists y [\mathsf{NTPP}(y,x)] \qquad \textbf{(NTPP)}$$

This axiom differs from those involving the quasi-Boolean functions in that it serves to introduce not only new regions but completely new boundaries. Clearly such *carve-up* axioms are needed if we want to be able to construct all possible figures by decomposing regions into parts in all possible ways that are distinguishable in terms of the theory. In the remainder of this section I look at some ways in which additional boundaries may be introduced into a figure.

5.1 Simple Building Blocks

We consider the possibilities for introducing new regions into an initial configuration made up of a number of designated ($2D$) regions which are externally connected and topologically simple: i.e. they are WCON. Assuming topological simplicity of the regions involved greatly simplifies the intellectual manageability of the problem. It means that a diagram representing some configuration of regions can be regarded as representative of all topologically equivalent diagrams. If we allowed that regions might be disconnected (or have some other topological complexity), then an example diagram would have to be thought of as also representing topologically distinct situations in which one or more of the designated regions were split into multiple parts. This would greatly reduce the usefulness of analysing diagrams as a means to finding existential axioms. Moreover, restricting our attention to complexes of topologically simple regions does not reduce the generality of the existential axioms, since regions with more complex topologies can be regarded as constructed, by means of sums and complements, from a number of simple regions; and the existence of such sums and complements would be guaranteed by axioms **QBfun1** and **QBfun2**.

5.2 Carving-Up a Simple Region

Starting with a single simple region there are three ways in which it can be divided into two self-connected parts — as shown in figure 3.

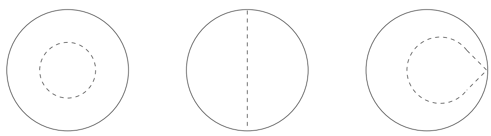

Fig. 3. Three intrinsically different ways of carving up a region

The possibility of the first case is guaranteed by the **NTPP** axiom in the RCC theory, however the other two possible carve-ups are not. The fact that the theory allows models in which there are regions with no tangential proper parts may be considered a shortcoming of the axioms. This could be remedied by adding the following additional axiom:

$$\forall x \exists y [\mathsf{TPP}(y, x)] \qquad \textbf{(TPP)}$$

As it stands **TPP** does not distinguish between the second and third cases shown in fig. 3; but such a distinction could be made by means of the 'firm tangential part' (FTPP) relation defined in terms of C by Gotts (1994b).

One idea (suggested by Dr A.G. Cohn) is that a full set of existential axioms might be arrived at by considering successive carve-ups starting from a single undifferentiated region. Thus, for each of the three cases of fig. 3 we would consider all the ways in which an additional division (into two self-connected parts) of one of the regions could be made. We have investigated each of the first two cases but space does not permit description of both, so I present only the further analysis of the middle case — see figure 4.

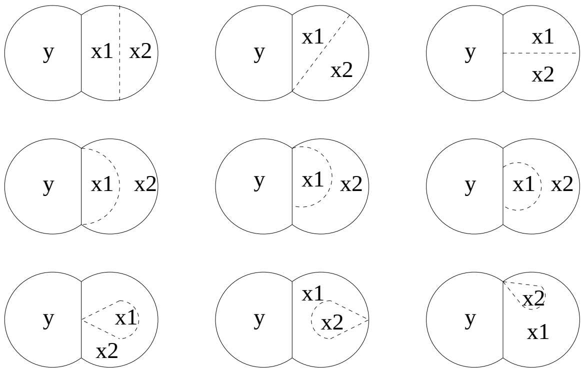

Fig. 4. Nine ways to split one of a pair of WCON, EC regions into two WCON parts

Each of the nine cases corresponds to an existential axiom which can be stated relatively easily in the RCC system. It will be helpful to define first the condition of two regions being each topologically simple and being externally connected along a boundary segment:

$$\mathsf{WCONFEC}(x, y) \ \equiv_{\text{def}} \ (\mathsf{WCON}(x) \wedge \mathsf{WCON}(y) \wedge \mathsf{FEC}(x, y))$$

and the relation which holds between a region and two well-connected parts into which it is split:

$$\mathsf{SPLIT}(x, y, z) \ \equiv_{\text{def}} \ (x = \mathsf{sum}(y, z) \wedge \neg\mathsf{O}(y, z) \wedge \mathsf{WCON}(y) \wedge \mathsf{WCON}(z))$$

These relations allow fairly concise specification of existential axioms for each situation. For example the top left case corresponds to the following:

$$\forall x \forall y [\mathsf{WCONFEC}(x, y) \rightarrow \exists x1 \exists x2 [\mathsf{SPLIT}(x, x1, x2) \wedge \mathsf{EC}(x1, y) \wedge \mathsf{DC}(x2, y)]$$

and the top middle case to:

$$\forall x \forall y [\mathsf{WCONFEC}(x, y) \rightarrow \exists x1 \exists x2 [\mathsf{SPLIT}(x, x1, x2) \wedge \mathsf{WCON}(\mathsf{sum}(x1, y)) \\ \wedge \mathsf{EC}(x2, y) \wedge \neg\mathsf{WCON}(\mathsf{sum}(x2, y))] \ ]$$

In fact all nine cases can be distinguished by means of simple RCC relations and/or conditions of 'well-connectedness' or otherwise of certain sums of regions.

If this method of analysing successive decompositions is to provide a complete set of existential axioms, we need to be able to show that, once we have considered a sufficient number of cases, then any carve-up of a more complex configuration can be accounted for in terms of a carve-up of simpler sub-parts. This seems eminently plausible; however, if it is not the case, then the analysis of possible bisections given in the next section may help complete the picture.

5.3 A Boundary Based Analysis

An alternative approach to specifying how additional boundaries my be introduced into arbitrarily complex configurations of (topologically simple) regions is to note that: whenever we split a region into two tangential parts, it is only the location of the end-points of the new bisecting line, relative to surrounding regions, that determines the topology of the resulting configuration.

E.g., if a region A is externally connected to two regions B and C then A can be carved into two pieces A_1 and A_2 in such a way that A_1 and A_2 are both connected to B and to C (see figure 5a). In other words A can be split from one boundary section to another, where these boundary sections are defined by external connection to some other region. Alternatively we may want to carve a region up in an even more specific way. If a region is externally connected to two other regions, such that all three regions are mutually connected at a point, then we may want to bisect the region at exactly that boundary point (see figure 5b). The other end of the bisection may be either at another such point (case c — not illustrated) or on a section of boundary defined by an external connection (case b).

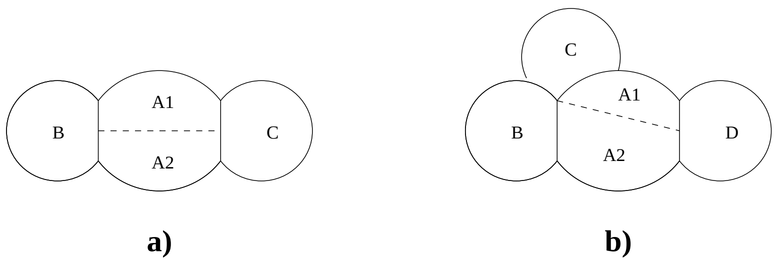

Fig. 5. Two ways to carve up a region

The existence of regions as constructed in case a) is assured by the following axiom:

$$\forall x \forall y \forall z[(\mathsf{WCON}(x) \wedge \mathsf{FEC}(x,y) \wedge \mathsf{FEC}(x,z)) \rightarrow \exists x1 \exists x2[\mathsf{SPLIT}(x,x1,x2) \wedge \mathsf{FEC}(x1,y) \wedge \mathsf{FEC}(x1,z) \wedge \mathsf{FEC}(x2,y) \wedge \mathsf{FEC}(x2,z)]]$$

and similar but slightly more complex formulae can easily be constructed to take account of cases b) and c).

This approach may obviate the need for the exhaustive case analysis proposed in the previous section. However it is not yet clear whether the boundary introductions together with the **NTPP** axiom suffice to introduce all possible regions. This is because in order to be able to describe bisections in terms of end-points, we need to first be able to generate configurations as complex as the situations shown in fig. 5 prior to dissection. Nevertheless, once configurations of sufficient complexity have been generated, the consideration of possible end-points for further bisections may be a better way to ensure completeness of the existential import of an axiom set.

5.4 How to Create the Universe

So far we have looked at how to construct new regions from the boundaries of other regions and how to carve regions up to produce new regions with new boundaries. But these forms of region generation will not get us very far unless we have some regions to combine or to chop up. The RCC theory ensures that there is a universal region, which is connected to all regions:

$$\mathsf{us} \;=_{\mathrm{def}}\; \iota y[\forall z[\mathsf{C}(z, y)]] \qquad \textbf{(us)}$$

We must consider whether starting from such a region and applying the (relative) existential axioms (which generate regions by combination and dissection of regions whose existence has already been secured) guarantees the existence of all possible configurations of regions. An answer to this question is beyond the scope of the present paper; however, a number of issues concerning the topology of the whole universe and how this relates to problems of defining topological properties of regions within the universe have been examined in (Gotts 1994b) and (Gotts 1994a).

5.5 Open Questions

The present work is only a preliminary enquiry into the possibility into how one might construct a complete theory of spatial regions. The examination of Boolean functions, dimension, planarity and carve up axioms given above cover a number of important aspects of the task but a number of open questions remain.

The axioms given in the last section are intended to completely specify possible configurations of regions. It will be noted that they are all 'positive': they ensure the existence of regions but do not explicitly rule out the existence of any regions. Nevertheless, because of the universal axioms of the RCC theory (which could themselves be regarded as negative existential axioms) and also by means of the dimensionality and planarity axioms suggested above, many possibilities are ruled out. However, we cannot yet be sure that all impossible combinations have been excluded; or, looking at the situation more dispassionately, we still may have several non-equivalent models of the axioms, so the theory may not be complete. This presents us with the question of whether fixing the global topology of a space and the dimensionality of the regions and then giving sufficient existential axioms to generate all possible configurations of regions will necessarily result in a complete theory.

Another problematic issue concerns the notions of the continuity of space and denumerability of regions. A categorical point-based theory of Euclidean space will contain some form of axiom of continuity. This is a 2nd-order formula (or infinite set of 1st-order formulae) which ensures that between any two non-overlapping open sets of points on a line there is always an intermediate point.

This has the consequence that there are must be an uncountable number of points in the domain. Hence, if regions are thought of as sets of points, one might expect that the domain of regions would be uncountable. However, a 1st-order theory, such as RCC plus the additional axioms suggested in this paper, will always have denumerable models. What is unclear is whether a satisfactory region-based theory really needs some additional 2nd-order axiom (akin to the continuity axiom) or whether a 1st-order theory with a unique denumerable model is perfectly adequate.

6 Conclusion

The lack of complete theories of spatial regions has been noted and attributed to certain weaknesses of current theories. Specifically, i) the dimension and global topology of the space under consideration has not been fixed axiomatically, and ii) insufficient attention has been given to to existential axioms.

To remedy i) I have given an axioms intended to characterise $2D$ planar space purely in terms of the connectedness of regions. To address ii) I have described a method of eliciting existential axioms by examining and formalising the process of construction of figures illustrating configurations of regions (a method analogous to the derivation of axioms of elementary geometry from a consideration of possible constructions with ruler and compass.)

Substantial work remains to be done before one could claim that an absolutely complete theory of spatial regions had been constructed. As yet it is not clear what form a proof of completeness should take. One could either take a semantic approach in which the theory was somehow interpreted in terms a unique mathematical structure (such as a Cartesian field) or a syntactic approach in which one attempted to directly show that for every formula either it or its negation were provable. It is also quite likely that rendering the theory complete will require further modification of the additional axioms I have suggested or even axioms of a kind which I have not anticipated. However, I believe that an absolutely complete theory of spatial regions will eventually be constructed.

Furthermore, the methodological considerations addressed in the paper are not only relevant to spatial reasoning but are important to the construction of formal theories describing almost any domain of knowledge. Whenever a set of concepts have a unique intended interpretation, an absolutely complete theory should be sought.

References

Allen, J. F.: 1981, An interval-based representation of temporal knowledge, *Proceedings 7th IJCAI*, pp. 221–226.

Asher, N. and Vieu, L.: 1995, Toward a geometry of common sense: A semantics and a complete axiomatization of mereotopology, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal.*

Bennett, B.: 1994, Spatial reasoning with propositional logics, *in* J. Doyle, E. Sandewall and P. Torasso (eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the 4th International Conference (KR94)*, Morgan Kaufmann, San Francisco, CA.

Biacino, L. and Gerla, G.: 1991, Connection structures, *Notre Dame Journal of Formal Logic*.

Clarke, B. L.: 1981, A calculus of individuals based on 'connection', *Notre Dame Journal of Formal Logic* **23**(3), 204–218.

Clarke, B. L.: 1985, Individuals and points, *Notre Dame Journal of Formal Logic* **26**(1), 61–75.

Cohn, A. G.: 1987, A more expressive formulation of many sorted logic, *Journal of Automated Reasoning* **3**, 113–200.

de Laguna, T.: 1922, Point, line and surface as sets of solids, *The Journal of Philosophy* **19**, 449–461.

Gotts, N. M.: 1994a, Defining a 'doughnut' made difficult, *in* C. Eschenbach, C. Habel and B. Smith (eds), *Topological Foundations of Cognitive Science*, Vol. 37 of *Reports of the Doctoral programme in Cognitive Science*, University of Hamburg.

Gotts, N. M.: 1994b, How far can we 'C'? defining a 'doughnut' using connection alone, *in* J. Doyle, E. Sandewall and P. Torasso (eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the 4th International Conference (KR94)*, Morgan Kaufmann.

Kuratowski, K.: 1972, *Introduction to Set Theory and Topology*, 2nd edn, Pergamon Press.

Ladkin, P.: 1987, *The Logic of Time Representation*, PhD thesis, University of California, Berkeley. Kestrel Institute report KES.U.87.13.

Leonard, H. S. and Goodman, N.: 1940, The calculus of individuals and its uses, *Journal of Symbolic Logic* **5**, 45–55.

Nebel, B.: 1995, Computational properties of qualitative spatial reasoning: First results, *Procedings of the 19th German AI Conference*.

Randell, D. A., Cui, Z. and Cohn, A. G.: 1992, A spatial logic based on regions and connection, *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, pp. 165–176.

Randell, D. and Cohn, A.: 1989, Modelling topological and metrical properties of physical processes, *in* R. Brachman, H. Levesque and R. Reiter (eds), *Proceedings 1st International Conference on the Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, Los Altos, pp. 55–66.

Scott, D.: 1959, Dimension in elementary euclidean geometry, *in* L. Brouwer, E. Beth and A. Heyting (eds), *The Axiomatic Method (with special reference to geometry and physics)*, North-Holland, Amsterdam, pp. 16–29.

Tarski, A.: 1956, Foundations of the geometry of solids, *Logic, Semantics, Metamathematics*, Oxford Clarendon Press, chapter 2. trans. J.H. Woodger.

Tarski, A.: 1959, What is elementary geometry?, *in* L. Brouwer, E. Beth and A. Heyting (eds), *The Axiomatic Method (with special reference to geometry and physics)*, North-Holland, Amsterdam, pp. 16–29.

Whitehead, A. N.: 1929, *Process and Reality*, The MacMillan Company, New York.