*Fundamenta Informaticae 34 (2000) 1–31*
*IOS Press*


# Empiricism and Rationalism in Region-based Theories of Space*

**Ian Pratt-Hartmann**
*Department of Computer Science*
*University of Manchester*
*Manchester, UK*
ipratt@cs.man.ac.uk

**Abstract.** Suppose we want to develop a theory of space in which the primary entities are not points, but *regions*. How do we proceed? Historically, the most popular approach is to select a group of spatial relations corresponding to familiar spatial concepts, and then to construct an axiomatic system governing these relations. The appropriateness of this axiomatic system is to be judged on the basis of its ability to chime with pre-theoretic intuition and of its success in a larger theory of scientific or commonsense spatial reasoning. In this paper, we argue that this approach is flawed, and leads only to labyrinthine systems of doubtful theoretical or practical value. We present an alternative approach which substitutes mathematical rigour for pre-theoretic intuition, and which at the same time provides a real guarantee of practical applicability.

**Keywords:** Spatial reasoning, mereology, topology, ontology

## 1. Introduction

Suppose we want to develop a theory of space in which the primary entities are not points, but *regions*. How do we proceed? Historically, the most popular approach is as follows. First, select a group of primitive spatial relations corresponding to familiar spatial concepts. Next, write down various axioms to govern these primitives. Finally, propose various definitions for a range of familiar spatial relations not included among the chosen set of primitives. The appropriateness of these axioms and definitions is to be judged on the basis of their ability to chime with pre-theoretic intuition and, perhaps, of their success in a larger theory of scientific or commonsense

*The author wishes to thank the referee for his valuable comments.

spatial reasoning. In particular, no attempt is made to *derive* or *deduce* the axioms of the theory, because there is nothing to derive or deduce them from.

We might call this approach, for want of a convenient label, the *empiricist* approach. Its radical character lies in the fact that region-based theories of space are constructed from the ground up, without any formal reference (except as a heuristic guide) to more familiar models of space based on the notion of points. Given this radicalism, the empiricism follows as a corollary: conformity to pre-theoretic intuition and participation in an empirically successful theory appear to be the only remaining criteria of evaluation for the proposed system of axioms and definitions.

In this paper, we argue that the approach just sketched is in fact a terrible way to develop region-based theories of space, and leads only to labyrinthine systems of axioms and definitions of doubtful theoretical or practical value. We present an alternative approach which substitutes mathematical rigour for pre-theoretic intuition, and which at the same time provides a real guarantee of practical applicability. On this alternative approach, region-based theories of space are constructed in terms of their relationships to more familiar, point-based models. This alternative approach is conservative and rationalist: conservative, because no attempt is made to build systems of axioms and definitions from the ground up; rationalist, because the appropriateness of the resulting region-based theories is secured by means of their logical relations to point-based models whose usefulness as representations of the space we inhabit is anyway beyond doubt. I will show that this approach leads to a clearer, more systematic understanding of region-based theories of space, full of surprising and non-trivial mathematical insights. To be sure, in contrasting "empiricist" and "rationalist" approaches, a certain simplification of positions is inevitable; and one runs the risk of unnecessarily raising hackles where there is in fact a substantial measure of agreement. But the labels are handy, and conveniently express the change in orientation for which I am arguing. Put simply: if you want to found a region-based theory of space, derive it from a familiar point-based model; do not try to hack the axioms from scratch.

## 2. The empiricist approach

The empiricist approach, as I have called it, is undoubtedly the method pursued by the pioneers of region-based theories of space, most notably, de Laguna [9] and Whitehead [24]. De Laguna employs as sole primitive a ternary relation which he calls "can-connect"; Whitehead, by contrast, substitutes the simpler binary relation "connects". For brevity of exposition, we will illustrate with examples from Whitehead's exposition, even though de Laguna's is clearly the better of the two. One terminological detail: in order to avoid confusion with the unrelated topological property of connectedness, we will use the word "contact" where Whitehead would have used "connection"; nothing of substance hinges on this change.

Intuitively, one region is in contact with another if and only if the two either overlap or touch at a point. Whitehead illustrates this notion of contact with the blobs drawn in figure 1. In familiar terms, if the blobs denote nonempty sets of points in the Euclidean plane, $A$ and $B$ are

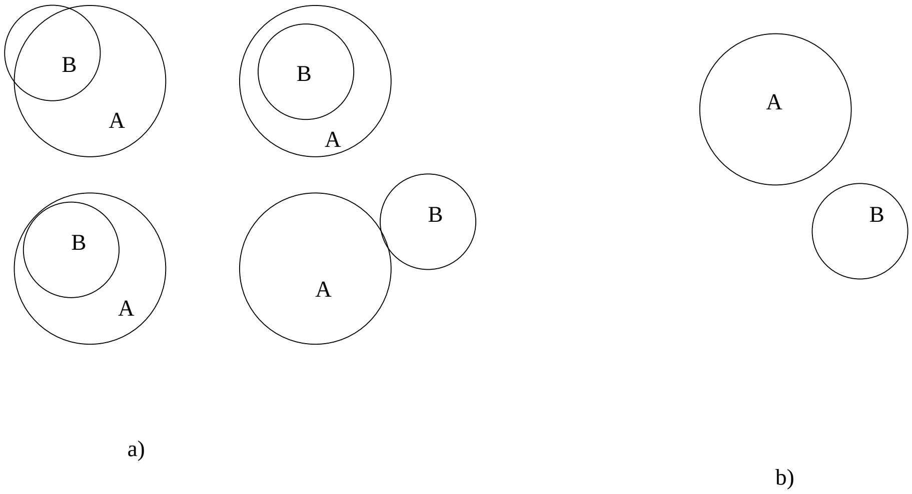



Figure 1 Whitehead's relation of contact ("connection"): in the four cases a), regions $A$ and $B$ are in contact; in the single case b), regions $A$ and $B$ are not in contact.

in contact if and only if their topological closures share a point. (Actually, this is not quite right: Whitehead bizarrely decrees that no region is in contact with itself; however, we will correct this oddity in the sequel and assume, as all subsequent authors have, that every nonempty region is in contact with itself.) From Whitehead's many axioms governing the notion of contact (he calls them "assumptions"), we choose the following two for the purposes of illustration:

> If $A$ is in contact with $B$, then $B$ is in contact with $A$.
>
> No region is in contact with every other region; and any two regions $A$ and $B$ are in contact with some third region $C$.

And from his many proposed definitions, we choose, again for illustration:

> Region $A$ is said to be a *part of* region $B$ if every region $C$ which is in contact with $A$ is also in contact with $B$.
>
> Regions $A$ and $B$ are said to *overlap* when there is some region $C$ which is a part of $A$ and a part of $B$.

And so it goes on: a long list. It is evident from these axioms and definitions that Whitehead assumes that there is no empty region and no universal region.

Clarke ([5] and [6]) presents a similar account of the contact relation (Clarke follows Whitehead in calling it "connection"), but couched in axiomatic terms using a formal language featuring a binary predicate $\mathsf{C}$. The underlying intuition is the same as in Whitehead's theory, except that (nonempty) regions are, sensibly, taken to be in contact with themselves. Clarke's system includes the following three axioms:

$$\begin{array}{l} \forall x \mathsf{C}(x,x) \\ \forall x \forall y (\mathsf{C}(x,y) \rightarrow \mathsf{C}(y,x)), \\ \forall x \forall y (\forall z (\mathsf{C}(x,z) \leftrightarrow \mathsf{C}(y,z)) \rightarrow x = y). \end{array}$$

Thus, contact is a reflexive and symmetric relation; and no two regions can contact exactly the same regions. The remaining axioms require a little more technical apparatus, and will not be given here.

As in Whitehead's theory, so in Clarke's, various mereotopological concepts are defined in terms of contact. For example, $\mathsf{P}(x,y)$ ($x$ is a part of $y$) and $\mathsf{O}(x,y)$ ($x$ overlaps $y$) are defined as follows:

$$\begin{array}{l} \mathsf{P}(x,y) := \forall z (\mathsf{C}(z,x) \rightarrow \mathsf{C}(z,y)) \\ \mathsf{O}(x,y) := \exists z (\mathsf{P}(z,x) \wedge \mathsf{P}(z,y)). \end{array}$$

Later accounts extend Clarke's strategy with new spatial primitives. Thus, for example, Randell, Cui and Cohn [21] add a function $\mathsf{conv\_hull}(x)$ denoting the convex hull of $x$, and augment the above axioms governing $\mathsf{C}$ with new axioms fixing the interpretation of $\mathsf{conv\_hull}(x)$. These axioms include, for example:

$$\begin{array}{l} \forall x \mathsf{P}(x, \mathsf{conv\_hull}(x)) \\ \forall x \mathsf{P}(\mathsf{conv\_hull}(\mathsf{conv\_hull}(x)), \mathsf{conv\_hull}(x)) \\ \forall x \forall y \forall z (\mathsf{P}(x, \mathsf{conv\_hull}(y)) \wedge \mathsf{P}(y, \mathsf{conv\_hull}(z)) \rightarrow \mathsf{P}(x, \mathsf{conv\_hull}(z))). \end{array}$$

(We have modified the notation in these axioms to avoid confusion with the *predicate* $\mathrm{conv}(x)$ introduced below.) In a similar vein, Borgo, Guarino and Masolo [3] introduce notions of congruence and sphericity into their region-based theory of space. The particular axioms and definitions they propose for this purpose are rather involved, and there is no need to repeat them here. The general approach of all these authors to the construction of region-based theories of space can be summed up as follows. First, select a group of primitive spatial relations corresponding to familiar spatial concepts. Next, write down various axioms to govern these primitives. Finally, propose various definitions for a range of familiar spatial relations not included among the chosen set of primitives.

Three distinct motivations for trying to develop a region-based theory of space recur in the literature. The first is the idea that regions are, in some sense, ontologically primary, because they correspond directly to the spaces occupied by mesoscopic objects; therefore, a theory of space should take regions as primitive entities and points as logical constructions based on them. This somewhat positivistic idea is most clearly articulated in Clarke [5] (p. 69 and p. 73), but similar sentiments can be found in virtually all papers presenting region-based theories of space.

A second motivation often given for region-based theories of space (and sometimes, for mereology more generally) is a preference for *elementary* theories: quantification over sets is thought undesirable, either on nominalistic grounds, or simply in order to make reasoning more tractable and easier to automate. Although Clarke makes extensive use of quantification over sets, more recent region-based theories of space, e.g. Gotts, Gooday and Cohn [11], stress the advantages from the point of view of AI of an elementary treatment. And while elementary point-based reconstructions of *geometry* have long been familiar (e.g. Tarski [23] and other papers in Henkin, Suppes and Tarski [12]), elementary reconstructions of *topology* seem bound to take regions as primitives.

A third motivation for region-based theories of space is more recent, and may be roughly characterized as a desire for *tameness*. The idea is this. Familiar, point-based models of space contain regions which could not possibly correspond to the three-dimensional space taken up by a physical object (or to the two-dimensional space taken up by its projection onto a surface). Fractal dust, odd collections of scattered points and the menagery of pathological constructions found in topology textbooks are examples. Likewise, familiar, point-based models of space permit distinctions which seem too fine-grained for everyday applications. The distinction between open, semi-open and closed regions is the most obvious example. By contrast, *region*-based theories of space, it is sometimes claimed, hold out the prospect of a tamed ontology from which pathological regions are banished and in which unwanted distinctions cannot be made. (See, for example, Randell *et al.* [21], p. 166, Borgo *at al.* [3], p. 221, Gotts *et al.* [11], sec. 3.).

This third motivation is particularly interesting from our point of view, because it suggests that the adoption of region-based account of space may involve a more radical shift than a mere tidying-up of our philosophical outlook, or a mere restriction of the logical resources at the disposal of automated spatial reasoning systems. Rather, the possibility arises that region-based theories of space may *conflict* (at least, apparently) with what familiar point-based models tell us about what regions exist. Suddenly, it seems, we have the freedom to invent whatever theory of space is demanded by the need to operate in and reason about our environment. Thus, for example, Randell *et al.* [21] contains a discussion of atomicity; or again, Asher and Vieu [1] considers a theory characterized by a class of models based on non-dense topological spaces.

However, perhaps the best motivation of all for investigating region-based theories of space is to be found in de Laguna's remark that such theories hold out the promise of an interesting, new perspective on an important phenomenon; and that is motivation enough. Consequently, we shall not concern ourselves further with motivational issues, though some of our conclusions may have an important bearing on them. Rather, our concern is with methodology: given that we want to develop a region-based theory of space, is the empiricist approach just sketched a sensible one?

The most obvious difficulty with the empiricist approach is the way the proposed axioms and definitions are evaluated. By "evaluation", here, we have nothing more definite in mind than whatever process is used to answer the questions that one asks when presented with any new theory. Most obviously: is the theory *true* as an an account of the space we inhabit? Why should we adopt just this particular set of axioms and not some other? Why is such-and-such a familiar spatial relation defined using just this formula and not some other? Less obviously, we might want to understand our theory's computational properties, its meta-theory and its relation to other theories. But it is the most basic questions of a theory's acceptability that we concentrate on here.

The authors cited above tend not to be very explicit about the justification for adopting one set of axioms and definitions rather than another. Clarke is perhaps the most forthcoming; and the closest he gets to a justification for his theory is the stress he lays ([5], p. 62) on the fact that the axioms he proposes, together with his reconstruction of points as sets of regions, have the consequence that two regions are in contact (satisfy the predicate $\mathsf{C}$) if and only if they share a common point. Thus, for Clarke, the ability of his theory to reproduce certain intuitions concerning relationships between points and regions is a key test of any region-based theory.

More generally, it seems that conformity to pre-theoretic intuition is the criterion implicitly appealed to in all of the region-based theories of space cited above. Consider the problem of deciding whether we should accept the axioms governing $\mathsf{C}$ and $\mathsf{conv\_hull}$ listed above. To the extent that we have a pre-theoretic grasp of the notions of contact and convex hull, we can ask ourselves whether the axioms in question *seem* reasonable. We can consider structures in which these axioms fail, and judge whether these structures would be ones we would countenance as possible models of space. If the answer to the first question is yes and the second is no, then that is a reason to accept the axioms. Likewise for the evaluation of proposed definitions: to the extent that we have a pre-theoretic grasp of both the definiendum and the primitives used in the definiens, we can ask ourselves whether the two notions thus picked out seem to coincide.

But reliance on intuitive reasonableness of axioms is a poor test of a region-based theory of space. There are two main drawbacks. First, while intuition may be of use in simple cases, there are, as we shall see, some highly non-obvious facts about familiar regions of space to whose (demonstrable) truth pre-theoretic intuition is a poor guide. Whitehead lists some thirty-one assumptions. Many of these assumptions employ defined terms, and so involve considerable unravelling. The axioms governing $\mathsf{C}$ in Randell *et al.* [21]—sometimes referred to as *RCC theory*—though fewer in number, often require similar unravelling. We should be skeptical as to whether intuition is able to deliver a clear verdict on such systems. Second, even to the extent that intuitive reasonableness is a useful test of the truth of any proposed list of axioms (interpreted as statements about the space we inhabit), it must remain silent on the question of whether any axioms have been omitted from that list. Thus, the theory of space presented may be half a theory of space: a rhapsodic assemblage of facts with—for all we know—many essential elements omitted. In fact, model-theoretic investigation of spatial logics paints a mixed picture in this regard. For example, an illuminating discussion of the model-theory of Clarke's axiom system is given in Biacino and Gerla [2]. Likewise, model-theoretic aspects of RCC theory are discussed in Lemon [14] and Renz [22]. But the one clear lesson we can draw from these investigations is that the axiom systems in question leave many formulas undecided.

It might be objected that the use of intuition to motivate postulates belongs to the process of discovery, not the process of justification. While our axioms may have been intuitively motivated, they are to be *experimentally* tested. For if our theory of space plays some role within a larger physical theory, or perhaps within a theory of commonsense reasoning, the adoption of a list of axioms might be justified if and insofar as the resulting overall theory successfully predicts observed phenomena or successfully guides goal-directed behaviour. Here again, however, the prospects for obtaining such a justification are bleak. The scope of the enquiry involved and the difficulty of assigning credit where theories perform well (and blame where they perform badly) seems insuperable. Thus, while such empirical considerations could justify a proposed axiomatization in principle, it is improbable that the requisite evidence could be marshalled in practice.

## 3. The rationalist alternative

This section introduces the rationalist approach to region-based theories of space advertised in the introduction. As with RCC theory, we shall be working within the framework of first-order logic, so that the primitive spatial relations are represented by predicate letters. In the sequel, we shall confine attention to the following predicate letters.

| | |
|---|---|
| $C(x, y)$ | region $x$ contacts region $y$ |
| $x \leq y$ | region $x$ is a part of region $y$ |
| $c(x)$ | region $x$ is connected |
| $\mathrm{conv}(x)$ | region $x$ is convex |

Selecting a collection of primitives then amounts to selecting a signature $\sigma$ for a first-order language. If $\sigma$ is any such signature, we denote the first-order language over $\sigma$ by $\mathcal{L}(\sigma)$. For example, we may elect to work within the language $\mathcal{L}(C)$, whose only primitive is $C(x, y)$, or perhaps within the language $\mathcal{L}(c, \leq)$, whose primitives are $c(x)$ and $x \leq y$.

Next, we take the familiar, point-based model of space—most obviously, $\mathbb{R}^2$ or $\mathbb{R}^3$—and identify regions with sets of these points in the normal way. Perhaps we do not want to allow *arbitrary* sets of points to qualify as regions, so we may select a subset $A$ of the set of all sets of points which count as *bona fide* regions, over which the variables of our region-based language will range. (We will have more to say below on the ramifications of making such a selection.) The spatial relations in the right-hand column of the above table are then defined over $A$ in the standard way. The definitions of *part-of*, *connected* and *convex* are so familiar that they do not need repeating here. As for *contacts*, the following definition has become standard:

**Definition 3.1.** Let $a$, $b$ be subsets of a topological space $X$. Then we say that *a contacts b* if the closures of $a$ and $b$ in $X$ have a point in common, i.e. if $a^- \cap b^- \neq \emptyset$.

These definitions give intuitively plausible reconstructions of the corresponding pre-theoretic notions. With the exception of the contact relation, they are firmly entrenched in the mathematical literature, and our accumulated mathematical knowledge of them fills whole libraries.

The result of the above choices is thus a precise specification of the set of allowed regions $A$, and a precise interpretation of each of the predicate letters in $\sigma$ as a set of tuples of elements from $A$. Such a construction is known in logic as a *structure*. We shall denote a structure obtained in this way by $A(\sigma)$. Since we always interpret the predicate letters standing for primitive spatial relations in the standard way, we can use $\sigma$ equivocally to denote both the predicate letters in our language and the relations over $A$ they denote. The important point about the structure $A(\sigma)$ is that it determines, via the usual semantics of first-order logic, the truth values of sentences of $\mathcal{L}(\sigma)$. (In the context of formal languages, a *sentence* is a formula with no free variables.) More generally, $A(\sigma)$ determines, for any tuple $\bar{a}$ of elements from $A$, which formulas are satisfied by $\bar{a}$. As usual in logic, we write $A(\sigma) \models \phi$ to say that a sentence $\phi$ is true in $A(\sigma)$, and $A(\sigma) \models \phi[\bar{a}]$ to say that $\bar{a}$ satisfies $\phi(\bar{x})$ in $A(\sigma)$. The set of all and only those sentences $\phi$ of $\mathcal{L}(\sigma)$ such that $A(\sigma) \models \phi$ is called the *theory* of $A(\sigma)$, written $\mathrm{Th}(A(\sigma))$.

Having chosen $\sigma$ and $A$, we regard the structure $A(\sigma)$ as our model of space—or as we shall say in the sequel, our *spatial ontology.* Thus, a spatial ontology is a model of what we think space is like at the level of regions: it tells us what regions exist and what properties those regions have. On the approach taken here, we begin with a (reasonably familiar) model of space, and *then* proceed to theorize about it.

A concrete example will help to make things clearer. Let our choice for the set of primitives be $\sigma = \{c, \leq\}$. Suppose we are working in two dimensions, so that our underlying point-based model of space is $\mathbb{R}^2$. Let $U$ be the set of all subsets of $\mathbb{R}^2$; and let our choice for the set of regions be $A = U$. (Thus, every collection of points, no matter how scattered or pathological, counts as a region.)

Now consider the sentence $\phi_{\mathrm{comp}}$ defined by

$$\begin{aligned}\phi_{\mathrm{comp}} := {} & \forall x \forall y (y \leq x \wedge c(y) \rightarrow \\ & \exists z (c(z) \wedge y \leq z \wedge z \leq x \wedge \forall w (c(w) \wedge z \leq w \wedge w \leq x \rightarrow w = z)).\end{aligned}$$

This sentence states that every connected subset of any region $x$ is contained within a component (a maximal connected subset) of $x$. Perhaps the most obvious question to ask about $\phi$ is: is it true? And, having chosen $U(c, \leq)$ as our spatial ontology, this question admits of a definitive answer. We have the following well-known result from elementary topology:

**Lemma 3.1.** *Let $a$ be a subsetset of a topological space $X$. Then every connected subset of a set $a$ is contained within a component of $a$.*

Hence,

**Observation 3.1.** $U(c, \leq) \models \phi_{\mathrm{comp}}$.

Of course, we cannot expect the truth of all sentences of $\mathcal{L}(c, \leq)$ in $U(c, \leq)$ to be so easily settled. But, as for $\phi_{\mathrm{comp}}$, so—at least in principle—for any other sentence: its truth value in a given structure is a matter to be settled by mathematical reasoning, not pre-theoretic intuition.

Given that $\phi_{\mathrm{comp}}$ is true in the chosen spatial ontology, $U(c, \leq)$, should $\phi_{\mathrm{comp}}$ be an *axiom*? That depends. It is perfectly sensible to ask whether the set of sentences $\mathrm{Th}(U(c, \leq))$ can be axiomatically characterized—that is, whether there is a (finitely presentable) system of axioms and rules of inference whose consequences are exactly $\mathrm{Th}(U(c, \leq))$. (To the best of the author's knowledge, this question has not been answered.) But if the answer is yes, then $\phi_{\mathrm{comp}}$ may or may not be one of those axioms: it depends on what happens to be the most convenient and succinct way of organizing the deductions. Again, notice that the question of whether a proposed axiomatic system is correct and complete for the set of sentences $\mathrm{Th}(U(c, \leq))$ is one of objective mathematical fact. That $\mathbb{R}^2$ is a sensible mathematical model of flat surfaces is not seriously in doubt; that the standard interpretations of the predicate letters $\leq$ and $c$ over $U$ conform well to our pre-theoretical notions of the part-whole relation and the property of connectedness is likewise firmly established. No further guarantee of intuitive reasonableness or practical applicability is required.

Having fixed our spatial ontology to be $U(c, \leq)$, our investigations need not stop at the axiomatic characterization of $\mathrm{Th}(U(c, \leq))$. Here are some of the more salient questions we might ask. What other strcutures make the set of sentences $\mathrm{Th}(U(c, \leq))$ true? What is their relationship to $U(c, \leq)$ in purely structural terms? Are any of these structures simpler than $U(c, \leq)$? Are there any other (more restrictive) choices for the set $A$ of regions which make the same sentences true? Do any other topological spaces lead to the same theory? What are the computational properties of $\mathrm{Th}(U(c, \leq))$? Is it decidable? What familiar spatial relations (reconstructed in the standard way over $U$) are *definable* as formulas of $\mathcal{L}(c, \leq)$?

These, then, are the questions that arise on the approach to region based-theories of space proposed here, which we may call the *rationalist* approach. To summarize, the approach is as follows. First, select a group $\sigma$ of predicate letters to denote various familiar spatial concepts. Next, we take the familiar, point-based model of space—most obviously, $\mathbb{R}^2$ or $\mathbb{R}^3$—and select a collection $A$ of subsets of this space to form the regions of the theory. Finally, interpret the predicate letters in $\sigma$ over $A$ in the standard way, yielding a structure $A(\sigma)$. It is this structure that is our proper subject of study.

At this point, some readers may be wondering whether this approach to region-based theories of space is in fact region-based at all. True, the variables in $\mathcal{L}(\sigma)$ range over elements of $A$, and hence over regions, not points. But these regions are identified with sets of points—indeed must be if the spatial primitives in $\sigma$ are to be interpreted in the ways that are familiar to us from mathematics. And this identification—so goes the objection—is exactly what the region-based approach is supposed to avoid.

The answer is that, in regarding $A(\sigma)$ as a *structure*, we view it in a way which abstracts away from the process by which it is constructed. This is most obvious when we address the issue of which other structures make $\mathrm{Th}(A(\sigma))$ true, and what their relationship is to $A(\sigma)$: if the region-based credentials of the original structure $A(\sigma)$ are open to doubt, those of an *arbitrary* model of $\mathrm{Th}(A(\sigma))$ are surely not. Of course, because *we* know how $A(\sigma)$ has been constructed, *we* can use that knowledge to answer questions about its theory, and hence about the class of structures sharing that theory. And it goes without saying that any technique from traditional geometry, topology and logic constitutes a legitimate tool for these purposes. Nevertheless, the conclusions we reach are in the first instance conclusions about structures of regions forming models of space which accurately reflect our everyday experience.

The remainder of this paper presents a survey of some of the salient questions regarding these structures, and the mathematical techniques required to answer them. Its aim is less to communicate the technical details of these results than to convey to the reader a flavour of the insights that they offer us—insights that would have been scarcely conceivable on the empiricist approach. It thus constitutes an extended advertisement for the rationalist approach to developing region-based theories of space.

# 4. Choosing the set of regions

The key decision to be made on the rationalist approach to region-based theories of space, apart from the choice of primitive spatial relations, is the selection of the set $A$ of regions. So far, we have looked only at the case $A = U$, but most region-based theories of space implicitly assume that we should be selective as to what counts as a region. One motivation for such selectivity is the following. Consider a physical object in three-dimensional space, such as a cup, or a natural two-dimensional projection of a physical object on a two-dimensional space, such as the land taken up by a building. Suppose we ask the question: does this region contain any (or all) of its boundary points? In topological terms: is the region open, semi-open or closed? The question seems meaningless. It cannot be resolved by a physical test, since the notion of the space occupied by a physical object makes no sense on very small scales. Neither can it be resolved by appeal to our commonsense conception of objects occupying definite regions of space, since that conception does not appear to take account of the distinctions between open and closed regions.

## 4.1. The regular open ontology

One way to avoid this awkward question is to restrict the set $A$ of regions so that *only* open sets are included. To make such a stipulation work elegantly, it has become common to restrict $A$ to the *regular open* sets:

**Definition 4.1.** Let $X$ be a topological space and $x \subseteq X$. Then the set $\bigcup\{y \subseteq X | y \text{ open}, y \cap x = \emptyset\}$ is an open set in $X$ called the *pseudocomplement of* $x$, written $x'$. We say that $x \subseteq X$ is *regular* if $x = x''$.

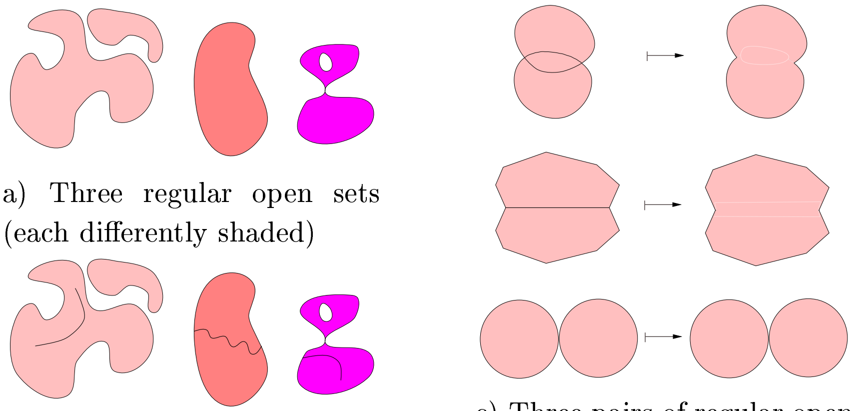
Thus, the pseudocomplement of a set $x$ is, in some sense, the 'nearest' thing we can have to the complement of $x$ if we restrict attention to open sets. Here is an alternative characterization of the regular open sets:

**Definition 4.2.** If $x$ is any set in a topological space, let $x^-$ denote the closure of $x$ and $(x)^0$ the interior of $x$. We write $\mathcal{F}(x)$ to denote the *frontier of* $x$, namely $x^- \setminus (x^-)^0$.

**Lemma 4.1.** *Let $x$ be a subset of a topological space $X$. Then $x' = X \setminus x^-$ and $x'' = (x^-)^0$. Hence, $x$ is regular if and only if $x = (x^-)^0$.*

The following well-known theorem underlies the importance of the regular open sets to mereotopology. We state it here without proof. (See, e.g. Johnstone [13], chapter I, section 1.13.)

**Theorem 4.1.** *Let $X$ be a topological space. Then the set of regular open sets in $X$ forms a Boolean algebra $\mathrm{RO}(X)$ with top and bottom defined by $1 = X$ and $0 = \emptyset$, and Boolean operations defined by $x.y = x \cap y$, $x + y = (x \cup y)''$ and $-x = x'$.*



a) Three regular open sets (each differently shaded)

b) Three nonregular open sets

c) Three pairs of regular open sets and their sums

Figure 2. Some regular open sets of the plane and their Boolean combinations

Basically, we can think of regular open sets in $\mathbb{R}^2$ as open sets with no internal cracks or point-holes (compare fig. 2a with fig. 2b). The product, $x.y$, of two regular open sets $x$ and $y$ is simply their intersection, which is guaranteed to be a regular open set. The sum, $x + y$, of two regular open sets $x$ and $y$ is a little more complicated; *very* roughly, it is the union of $x$ and $y$ with any internal boundaries removed (fig. 2c). Finally, the pseudocomplement, $-x$, of a regular open set $x$ is simply that part of the plane not occupied by $x$ or its boundary.

Let us denote by $O$ the set of regular open sets of the plane. Then $O(c, \leq)$ forms a structure for our spatial language in the usual way. The interpretations of the relations $c$ and $\leq$ will be as before, except that they will be restricted now to the smaller domain of interpretation.

A word about the status of the function symbols $+$, $.$ and $-$, and the constants 1 and 0. These can be defined in the standard way for *any* collection $A$ of regions forming a Boolean algebra under the relation $\leq$. Moreover, since these definitions are expressible in first-order terms, we may use the function symbols $+$, $.$, $-$ and the constants 1, 0 freely in formulas of $\mathcal{L}(\sigma)$, provided $\sigma$ contains the predicate letter $\leq$, understanding them as abbreviations. This simply makes formulas more readable, and we shall make frequent use of this device in the sequel. Beware, however, that the values of these functions may be different in different structures. For example, if $a$ and $b$ are regions in $O$ (and hence in $U$), their sum in $U$ is their union; however, their sum in $O$ may be a strict superset of their union. One further abbreviation: we use $x < y$ as a shorthand for for $x \leq y \wedge x \neq y$.

Changing the set of regions $A$ over which our variables range will in general change the set of sentences $\text{Th}(A(\sigma))$. For example, the set $U$ contains singleton sets, which are *atoms*—non-empty regions with no non-empty proper parts. By contrast, all nonempty sets in $O$ contain proper parts. Note that, in both $U$ and $O$, the constant 0 denotes the empty set. Hence if $\phi_{\text{nonat}}$ is the sentence

$$\phi_{\text{nonat}} := \forall x(0 < x \rightarrow \exists y(y \neq 0 \wedge y < x)),$$

then we have $O(\leq) \models \phi_{\text{nonat}}$ but $U(\leq) \not\models \phi_{\text{nonat}}$.

**Definition 4.3.** Given two models $\mathfrak{A}$ and $\mathfrak{B}$, we say that $\mathfrak{A}$ and $\mathfrak{B}$ are *elementarily equivalent*, written $\mathfrak{A} \equiv \mathfrak{B}$, if $\text{Th}(\mathfrak{A}) = \text{Th}(\mathfrak{B})$.

Hence,

**Observation 4.1.** If $\sigma$ contains the predicate letter $\leq$, then $U(\sigma) \not\equiv O(\sigma)$.

It is worth pointing out that many sentences of $\text{Th}(U(c, \leq))$ remain true when we restrict attention to $O$. Recall the $\mathcal{L}(c, \leq)$-sentence $\phi_{\text{comp}}$ defined above. The following result is standard:

**Lemma 4.2.** *Every component (maximal connected subset) of a regular open set in some topological space is regular open.*

Hence:

**Observation 4.2.** $O(c, \leq) \models \phi_{\text{comp}}$.

## 4.2. The polygonal ontology

From the point of view of eliminating space-junk, the regular open sets $O$ constitute an improvement over $U$ in that restricting quantification to $O$ suppresses distinctions which depend only on boundary points. Nevertheless, lurking in $O$ are sets of points which could never correspond to the spaces occupied by physical objects or their two-dimensional projections.

Consider, for example, the regions $a_1$, $a_2$ and $a_3$ defined by

$$\begin{aligned} a_1 &= \{(x, y) | -1 < x < 0\ ;\ -1 - x < y < 1 + x\} \\ a_2 &= \{(x, y) | 0 < x < 1\ ;\ -1 - x < y < \sin(1/x)\} \\ a_3 &= \{(x, y) | 0 < x < 1\ ;\ \sin(1/x) < y < 1 + x\}\ , \end{aligned}$$

and depicted in fig. 3. It is easy to see that $a_1$, $a_2$ and $a_3$ are all regular open sets. Region $a_1$ is unexceptionable: it is just a triangle. But $a_2$ and $a_3$ seem to be the sorts of regions we might want to to exclude from the set $A$ which our region-based theory of space quantifies over.

The simplest way to do this is to take $A$ to be the set of regular open *polygons*. (In fact, such a severe restriction is not needed for almost all of the results that follow; however, we shall concentrate on polygons for simplicity.) Any line in the plane cuts the plane into two connected, open sets, called *half-planes*. It is easy to see that these sets are regular, with each being the pseudocomplement of the other. Hence, we can speak about their sums, products and complements in $\text{RO}(\mathbb{R}^2)$.

**Definition 4.4.** A *basic polygon in* $\mathbb{R}^2$ is the intersection of finitely many half-planes in $\mathbb{R}^2$. A *polygon in* $\mathbb{R}^2$ is the sum, in the Boolean algebra $\text{RO}(\mathbb{R}^2)$, of any finite set of basic polygons in $\mathbb{R}^2$.

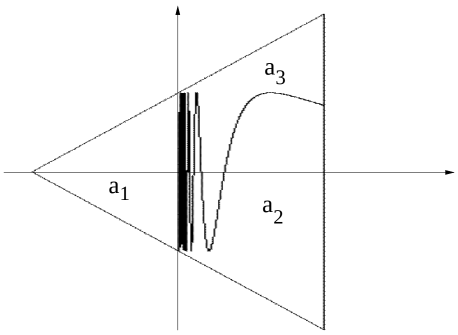



Figure 3. Some wiggly regions

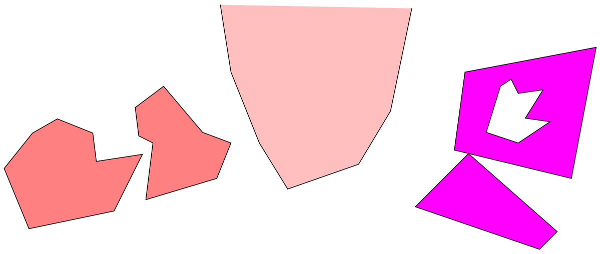


Figure 4. Some polygons

We denote the set of polygons in $\mathbb{R}^2$ by $R$. Fig. 4 shows some polygons.

Note that polygons, in this sense, need not be connected and need not be bounded. Note also that $\emptyset$ and $\mathbb{R}^2$ count as polygons. We have the following result:

**Lemma 4.3.** *$R$ is a Boolean subalgebra of* $\mathrm{RO}(\mathbb{R}^2)$.

**Proof:**
We need only show that $R$ is closed under the Boolean operations. But this is obvious given the distribution laws for $\mathrm{RO}(\mathbb{R}^2)$. $\square$

Idealizing slightly, $R$ is the set of regions recognized by most computer systems specialized for handling plane spatial data, such as geographic information systems (*GIS*s). That is, such systems are limited to regions whose boundaries are made up of finitely many lines and line segments. Experience has shown that, for many practical purposes, no loss of useful modelling results from limiting attention to polygons. Thus, one sensible spatial ontology is provided by the structure $R(c, \leq)$; in the sequel, we will derive many surprising facts about the theory of this structure.

Again, establishing the truth of sentences interpreted over $R$ requires just as much care as for $O$. Of course, $R$ has no atoms, from which we immediately see that

**Observation 4.3.** $R(c, \leq) \not\equiv U(c, \leq)$.

Nevertheless, the sentence $\phi_{\text{comp}}$ remains true even when we restrict quantification to $R$. Formally:

**Lemma 4.4.** *Let $a \in R$ and let $b$ be a component of $a$. Then $b \in R$.*

It follows of course that $R(c, \leq) \models \phi_{\text{comp}}$.

So much for the relationship between $R(c, \leq)$ and $U(c, \leq)$. But what of the relationship between $R(c, \leq)$ and $O(c, \leq)$? Consider the sentence $\phi_{\text{con}}$ given by

$$\begin{aligned}\phi_{\text{con}} := \forall x_1 \forall x_2 \forall x_3 ((\bigwedge_{1 \leq i \leq 3} c(x_i) \wedge c(x_1 + x_2 + x_3)) \rightarrow \\ (c(x_1 + x_2) \vee c(x_1 + c_3))).\end{aligned}$$

Should we count this sentence as true? Well, it is not difficult to show

**Observation 4.4.** $R(c, \leq) \models \phi_{\text{con}}$.

However, this result does not generalize to the spatial ontology $U(c, \leq)$. For recall the regions $a_1$, $a_2$ and $a_3$ illustrated in figure 3. It is not difficult to show that $a_1 + a_2 + a_3$ is the interior of the large triangle in fig. 3, and that $a_1 + a_2 = a_1 \cup a_2$ and $a_1 + a_3 = a_1 \cup a_3$. Hence $a_1 + a_2 + a_3$ is connected, but $a_1 + a_2$ and $a_1 + a_3$ are not. It follows that $O(c, \leq) \not\models \phi_{\text{con}}$. Hence:

**Observation 4.5.** $O(c, \leq) \not\equiv R(c, \leq)$.

This case demonstrates well the difference in style and content between what we have been calling the empiricist and rationalist approaches to developing region-based theories of space. Should the sentence $\phi_{\text{con}}$ be accepted as a truth about the primitive spatial relations $\leq$ and $c$? For the rationalist, the answer is easy: it depends. If we mean to quantify only over polygonal regions, the answer is yes; if we mean to quantify over very wiggly regions, the answer is no. Of course, in some sense, this just pushes the question back: what regions *do* we mean to quantify over? But there seems to be no reason in principle why we should not be able to answer this question, given that our answer may depend on the application we have in mind. Certainly, in the light of the arrangement of figure 3, it is dangerous to consult our pre-theoretic intuitions as to the truth of $\phi_{\text{con}}$!

## 4.3. The rational polygonal ontology

So far, we have encountered three spatial ontologies: $U(c, \leq)$, $O(c, \leq)$ and $R(c, \leq)$—all having different theories. We now turn to a further restriction of the set of regions which makes *no* difference to the corresponding theory over the relations $c$ and $\leq$.

If a line is defined by an equation $ax + by + c = 0$, where $a$, $b$ and $c$ are rational numbers, we call it a *rational line*; and if a half-plane is bounded by a rational line, we call it a *rational half-plane*. It is obvious that a pair of rational lines can intersect only at points with rational coordinates. Now we define:

**Definition 4.5.** A *rational basic polygon* is the intersection of finitely many rational half-planes. A *rational polygon* is the sum, in the Boolean algebra $\mathrm{RO}(\mathbb{R}^2)$, of any finite set of rational basic polygons.

We denote the set of rational polygons by $Q$, and will sometimes refer to it as the *rational polygonal domain.* In fact, $Q$ may have a better claim to be a suitable ontology for commonsense spatial reasoning than $R$. As before, interpreting the primitives $c$ and $\leq$ over $Q$ in the normal way gives us the structure $Q(c, \leq)$.

The question naturally arises as to the relationship between $Q(c, \leq)$ and $R(c, \leq)$. To understand this relationship, we need some simple topology and some simple model theory. First, the simple topology.

**Definition 4.6.** Let $X$ be a topological space and $a_1, \ldots, a_n$, $b_1, \ldots, b_n$ subsets of $X$. We say that $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are *similarly situated*, and write $a_1, \ldots a_n, \sim b_1, \ldots, b_n$, if there is a homeomorphism of the space $X$ onto itself taking $a_i$ to $b_i$ for all $i$ $(1 \leq i \leq n)$.

**Lemma 4.5.** *Let $a_1, \ldots, a_n, b_1, \ldots, b_n, a \in R$ such that $a_1, \ldots, a_n \sim b_1, \ldots, b_n$. Then there exists $b \in R$ such that $a_1, \ldots, a_n, a \sim b_1, \ldots b_n, b$.*

**Lemma 4.6.** *Let $a_1, \ldots, a_n \in Q$ and $b \in R$. Then there exists $a \in Q$ such that $a_1, \ldots, a_n, b \sim a_1, \ldots, a_n, a$.*

We omit the proofs of these lemmas, which are standard and anyway intuitively obvious. The first has a simple but important corollary about the structure $R(c, \leq)$.

**Lemma 4.7.** *Let $\bar{a}$ and $\bar{b}$ be tuples in $R$ such that $\bar{a} \sim \bar{b}$. Then $\bar{a}$ and $\bar{b}$ satisfy the same formulas in $R(c, \leq)$.*

**Proof:**
By structural induction on the complexity of formulas, using lemma 4.5. □

Hence, the language $\mathcal{L}(c, \leq)$ is unable to distinguish between similarly situated tuples. Next, the simple model theory:

**Definition 4.7.** Let $\mathfrak{A}$ and $\mathfrak{B}$ be structures. We say that $\mathfrak{A}$ is an *elementary submodel* of $\mathfrak{B}$, written $\mathfrak{A} \preceq \mathfrak{B}$, if $A \subseteq B$ and for any tuple $\bar{a}$ from $A$ and any formula $\phi(\bar{x})$, $\mathfrak{A} \models \phi[\bar{a}]$ if and only if $\mathfrak{B} \models \phi[\bar{a}]$.

Thus, $\mathfrak{A}$ is an elementary submodel of $\mathfrak{B}$ if $\mathfrak{B}$ recognizes a collection of additional elements which make no difference to the first-order expressible properties possessed by the elements of $A$. Note that $\mathfrak{A} \preceq \mathfrak{B}$ immediately implies $\mathfrak{A} \equiv \mathfrak{B}$.

**Theorem 4.2. (Tarski-Vaught)** *Let $\mathfrak{A}$ and $\mathfrak{B}$ be structures such that $\mathfrak{A} \subseteq \mathfrak{B}$ and, for any $n$-tuple $\bar{a}$ of $\mathfrak{A}$ and any formula $\phi(\bar{x})$ of the form $\exists y \psi(\bar{x}, y)$ such that $\mathfrak{B} \models \phi[\bar{a}]$, there exists $a \in A$ such that $\mathfrak{B} \models \psi[\bar{a}, a]$. Then $\mathfrak{A} \preceq \mathfrak{B}$.*

Now we can state the relationship between $Q(c, \leq)$ and $R(c, \leq)$.

**Observation 4.6.** $Q(c, \leq) \preceq R(c, \leq)$.

**Proof:**
By construction, $Q(c,\leq) \subseteq R(c,\leq)$. Let $\bar{a}$ be an n-tuple of elements of $Q$, and let $\phi(\bar{x})$ be any formula of $\mathcal{L}$ of the form $\exists y \psi(\bar{x}, y)$ such that $R(c,\leq) \models \phi[\bar{a}]$. Then there exists $b \in R$ such that $R(c,\leq) \models \psi[\bar{a}, b]$. By lemma 4.6, there exists $a \in Q$ such that $\bar{a}, a \sim \bar{a}, b$. By lemma 4.7, $R(c,\leq) \models \psi[\bar{a}, a]$. The result then follows from theorem 4.2. □

We now have a good idea about the relationship between the spatial ontologies $Q(c,\leq)$ and $R(c,\leq)$. Observation 4.6 tells us that these structures have identical theories. Moreover, it is clear that they are essentially different, if only for the obvious reason that the former is countable and the latter uncountable. However, observation 4.6 also tells us that the extra regions recognized by $R(c,\leq)$—namely, the irrational polygons—make no difference to the $\mathcal{L}(c,\leq)$-expressible properties possessed by tuples of rational polygons. In some sense, then, $R(c,\leq)$ is an *unnecessarily complicated* ontology for the purposes of the language $\mathcal{L}(c,\leq)$.

## 4.4. A nontopological example

The previous result can be regarded as establishing something about the *inexpressiveness* of the language $\mathcal{L}(c,\leq)$: for it says that this language cannot distinguish between the ontologies $R$ and $Q$ in terms of the sentences the corresponding structures make true. Examination of the proof of observation 4.6 shows that the result generalizes to any signature $\sigma$ of purely topological relations. The question naturally arises as to whether the (non-topological) property of convexity allows us to do any better. The following observation shows us that it does. Recall that the predicate $\text{conv}(x)$ is always taken to be satisfied by a region $a$ just in case $a$ is convex.

**Observation 4.7.** The structures $Q(\text{conv},\leq)$ and $R(\text{conv},\leq)$ are not elementarily equivalent.

**Proof:**
A region in $R$ or $Q$ satisfying the formula

$$\text{conv}(x) \wedge x \neq 0 \wedge \text{conv}(-x) \wedge x \neq 1$$

is a half-plane, and so can be taken to stand for the line that bounds it. In the sequel, then, we shall say that a *line* satisfies a formula when we really mean that some half-plane standing for it does so. If the lines $k$, $l$, $m$ form a triangle, let us denote the intersection points $m \cap l$ and $k \cap l$ by $O$ and $I$, respectively. (See figure 5.) If $j_b$ is any line intersecting $l$ in a point $B$, it is easy to see that $B$ lies on the same side of $O$ as $I$ if both $k$ and $j_b$ lie entirely outside the same quadrant created by $l$ and $m$ (marked with $\star$ in the diagram). Hence it is easy to construct an $\mathcal{L}(\text{conv},\leq)$-formula $\text{pos}(u_b, u_k, u_l, u_m)$ satisfied by the lines $j_b$, $k$, $l$ and $m$ guaranteeing that $k$, $l$ and $m$ form a triangle with the points $l \cap j_b$ and $l \cap k$ lying on the same side of the point $l \cap m$ on $l$.

In figure 5, the lines $j_a$, $j_b$ and $j_c$ are shown intersecting $l$ at the points $A$, $B$ and $C$, respectively, and satisfy the conditions: $j_a || k$, $j_c \cap m = k \cap m = \{X\}$, $j_b || j_c$ and $j_b \cap m = j_a \cap m = \{Y\}$.

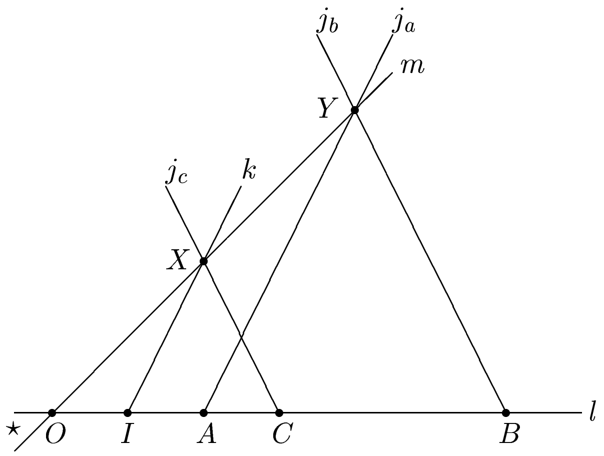



Figure 5. Multiplication of points on a line

By easy plane geometry, triangles $OXI$ and $OYA$ are similar, as indeed are triangles $OXC$ and $OYB$. Hence:

$$\begin{aligned}\overline{OY}/\overline{OX} &= \overline{OA}/\overline{OI} \\ \overline{OY}/\overline{OX} &= \overline{OB}/\overline{OC}.\end{aligned}$$

Therefore,

$$\overline{OA}/\overline{OI} = \overline{OB}/\overline{OC}.$$

whence

$$\frac{\overline{OA}}{\overline{OI}} \cdot \frac{\overline{OC}}{\overline{OI}} = \frac{\overline{OB}}{\overline{OI}}. \tag{1}$$

Intuitively, equation (1) says that, if $A$, $B$ and $C$ are taken to denote numbers on the line $l$, where $O$ is the origin and $\overline{OI}$ the unit of measurement, then $B = A.C$. Since the conditions on $j_a$, $j_b$ and $j_c$ are certainly expressible in $\mathcal{L}(\text{conv}, \leq)$, there is a formula $\text{mult}(u_a, u_b, u_c, u_k, u_l, u_m)$ satisfied by lines $j'_a, j'_b, j'_c, k, l, m$ just in case $k$, $l$ and $m$ form a triangle as depicted in figure 5, and $j'_a$, $j'_b$ and $j'_c$ intersect $l$ in points $A$, $B$ and $C$ respectively, such that equation (1) holds. (Note, however, that we do not require $j'_a$, $j'_b$ and $j'_c$ to be arranged as $j_a$, $j_b$ and $j_c$ in figure 5.)

Now let $\phi$ be the sentence

$$\forall u_b \forall u_k \forall u_l \forall u_m (\text{pos}(u_b, u_k, u_l, u_m) \rightarrow \exists u \text{mult}(u, u_b, u, u_k, u_l, u_m)).$$

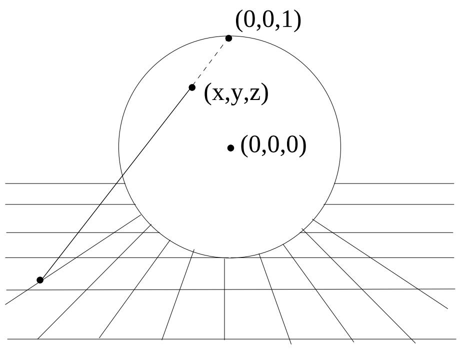



Figure 6. Stereographic projection of the unit sphere onto the closed plane

Then it is obvious that $R(\text{conv}, \leq) \models \phi$. For let $j_b$, $k$, $l$, $m$ satisfy $\text{pos}(u_b, u_k, u_l, u_m)$ in $R(\text{conv}, \leq)$, and let the points $O$, $I$, $B$ be as in figure 5. Then we can find a suitable witness for $u$ in $\phi$ by taking a line $j$ intersecting $l$ at a point $A$ where $\overline{OA}/\overline{OI} = \sqrt{\overline{OB}/\overline{OI}}$. That $Q(\text{conv}, \leq) \not\models \phi$ is equally obvious if we choose $j_b$ intersecting $l$ at a point $B$ such that $\overline{OB}/\overline{OI}$ has an irrational square root. □

Thus, over the signature $(\text{conv}, \leq)$, which region-based theory of space we adopt turns out to be sensitive to whether we take the underlying model of space to be the real plane or the rational plane. Which is the best theory for the purposes of commonsense reasoning? It is hard to say. But it is a strength of the perspective taken here that we at least have a clear idea of what the issue is and what it depends on.

### 4.5. Another underlying space

All the examples considered so far are based on the same topological space, $\mathbb{R}^2$. However, there is no reason spatial ontologies cannot be based on other familiar (or indeed unfamiliar) topological spaces. Of particular interest is the closed plane $\mathbb{R}^2 \cup \{\infty\}$, under the usual topology. In the sequel, we denote the closed plane by $Z^2$. It helps to visualize $Z^2$ as the surface of the sphere $x^2 + y^2 + z^2 = 1$ in $\mathbb{R}^3$, which can be stereographically projected onto the plane $z = -1$ via the mapping

$$(x, y, z) \quad \mapsto \quad \left(\frac{2x}{1-z}, \frac{2y}{1-z}, -1\right)$$

with the point $(0, 0, 1)$ going to the point at infinity, $\infty$. The projection is illustrated in fig. 6. In the closed plane, all lines are taken to pass through the point $\infty$. Under the stereographic projection, lines in the plane correspond to great circles on the surface of the sphere $x^2+y^2+z^2 = 1$, passing through the polar point $(0, 0, 1)$.

It is easy to show that every line divides $Z^2$ into two regular connected sets, which we shall again refer to as half-planes. Since lines contain the point $\infty$, half-planes in the closed plane are literally the same sets of points as half-planes in the open plane.

**Definition 4.8.** A *basic polygon in* $Z^2$ is the intersection of finitely many half-planes in $Z^2$. A *polygon in* $Z^2$ is the sum, in the Boolean algebra $\mathrm{RO}(Z^2)$, of any finite set of basic polygons in $Z^2$.

We denote the set of polygons in $Z^2$ by $\bar{R}$.

Again, it is easy to show that $\bar{R}$ is a Boolean subalgebra of $\mathrm{RO}(Z^2)$, so we define the closed-plane model $\bar{R}(c, \leq)$ by interpreting the primitives in the usual way, relative, of course, to the topology on $Z^2$. Denote the set of open sets in $Z^2$ (under the usual topology) by $\Omega$. Let $\hat{x} = x \setminus \{\infty\}$ for all $x \in \Omega$ and let $\hat{\Omega} = \{\hat{x} | x \in \Omega\}$. Then $\langle \mathbb{R}^2, \hat{\Omega} \rangle$ is the usual topology on the open, real plane. Moreover, $\hat{\Omega} \subseteq \Omega$. If $u \in \hat{\Omega}$, then the pseudocomplement $u^*$ of $u$ in $\hat{\Omega}$ is in general different to the pseudocomplement $u'$ of $u$ in $\Omega$; and the property of being regular in $\hat{\Omega}$ is different to the property of being regular in $\Omega$. However, we have the following results:

**Lemma 4.8.** *If* $x \in \Omega$*, then* $\hat{x}^* = \widehat{x'}$.

**Proof:**
Routine. □

**Lemma 4.9.** *If* $x \in \Omega$ *is regular in* $\Omega$*, then* $\hat{x}$ *is regular in* $\hat{\Omega}$.

**Proof:**
By two applications of lemma 4.8, $\hat{x}^{**} = \widehat{x''} = \hat{x}$. □

**Lemma 4.10.** *If* $u \in \hat{\Omega}$ *is regular in* $\hat{\Omega}$*, then there exists an* $x \in \Omega$ *regular in* $\Omega$ *such that* $\hat{x} = u$.

**Proof:**
Recall that $\hat{\Omega} \subseteq \Omega$. It is easy to show that $u \in \Omega$ implies that $u''$ is regular in $\Omega$. But again, by lemma 4.8, $\widehat{u''} = \hat{u}^{**} = u^{**} = u$. □

Thus, although $R(c, \leq)$ and $\bar{R}(c, \leq)$ are based on nonhomeomorphic topological spaces, they are very closely related, as the following result shows.

**Definition 4.9.** Given any two structures $\mathfrak{A}$ and $\mathfrak{B}$ interpreting the same signature, we say that $\mathfrak{A}$ and $\mathfrak{B}$ are *isomorphic*, written $\mathfrak{A} \simeq \mathfrak{B}$, if there is a 1–1 onto mapping $f : A \to B$ such that the image under $f$ of the interpretation of any primitive in $\mathfrak{A}$ is the interpretation of that primitive in $\mathfrak{B}$.

It is a standard result that, if $\mathfrak{A} \simeq \mathfrak{B}$, then $\mathfrak{A} \equiv \mathfrak{B}$.

**Observation 4.8.** The structures $R(c, \leq)$ and $\bar{R}(c, \leq)$ are isomorphic.

**Proof:**
Lemmas 4.9 and 4.10 show that $f : x \mapsto \hat{x}$ maps the set of regular sets in $\Omega$ onto the set of regular sets in $\hat{\Omega}$. And since no two regular sets in $\Omega$ differ by a single point, this mapping is 1–1. Moreover, given the fact that, for all $x, y \in \Omega$, $\hat{x} \cap \hat{y} = \widehat{x \cap y}$, lemma 4.8 further guarantees that $f$ induces a Boolean algebra isomorphism taking $\mathrm{RO}(Z^2)$ to $\mathrm{RO}(\mathbb{R}^2)$. And since $f$ maps half-planes in $Z^2$ to half-planes in $\mathbb{R}^2$, $f : \bar{R} \to R$ is also a Boolean algebra isomorphism. Finally, we observe that $x \in \Omega$ is connected in $\Omega$ if and only if $\hat{x}$ is connected in $\hat{\Omega}$, so that $f : \bar{R}(c, \leq) \to R(c, \leq)$ is a model isomorphism. □

It follows of course that $\mathrm{Th}(\bar{R}(c, \leq)) = \mathrm{Th}(R(c, \leq))$. In the sequel, it will sometimes be convenient to use $\tilde{R}$ ambiguously to denote either $R$ or $\bar{R}$.

### 4.6. Axiomatization

So far, we have made a series of rather specific observations comparing and contrasting the theories generated by different spatial ontologies. But we have not so far characterized any of these theories in its entirety. Given that the first attempts to develop region-based theories of space were presented as axiomatic systems, it seems reasonable to ask whether the rationalist approach presented here has anything interesting to say on this matter.

In fact, an axiomatization of $\mathrm{Th}(\tilde{R}(c,\leq)) = \mathrm{Th}(Q(c,\leq))$ is presented in Pratt and Schoop [17]. The axiomatization is somewhat complicated, and little would be gained by simply reproducing it here. We should, however, point out that this axiomatization is accompanied by a proof of its correctness and completeness. That is, it can be shown not only that the consequences of the axiom system are all true in the structures $\tilde{R}(c,\leq)$ but that, conversely, all truths of $\tilde{R}(c,\leq)$ are consequences of the axiom system. We note in passing that this axiomatization incorporates an *infinitary* proof rule. The need for some such non-standard device follows from the fact that membership of a formula in $\mathrm{Th}(\tilde{R}(c,\leq))$ is in fact undecidable, as was shown by Dornheim [10].

### 4.7. Summary

This section has presented a plethora of technical results on the relationships between choices of spatial ontology and the theories they give rise to. But in some sense, these are only of secondary importance. The real lesson of these investigations is that the trouble we took to provide detailed characterizations of our region-based spatial ontologies was worth the effort. For the details really make a difference: what is true for one ontology may be false for another.

Hence it is unclear what the empiricist approach is appealing to when it asks us to use our intuition to judge the reasonableness of proposed axioms and definitions. Should $\phi_{\mathrm{con}}$ be an axiom (or a consequence of our axioms)? The answer: it depends on how generous our spatial ontology is when it comes to doling out regionhood to sets of points in the plane. The rationalist approach gives us a principled criterion against which to judge a sentence such as $\phi_{\mathrm{con}}$, relative, of course, to a choice of spatial ontology. The empiricist approach, by contrast, gives us a blank choice.

## 5. Defining topological relations

In this section we turn to the question of the definition of familiar spatial relations in terms of a chosen set of primitives. As before, we stress the sensitivity of our results to choices of spatial ontology and the importance of rigorous mathematical argument.

### 5.1. Definability and non-definability in the language $\mathcal{L}(C)$

To illustrate the issues, consider again Whitehead's definition of the part-whole relation in terms of contact: $A$ is a part of $B$ just in case every region in contact with $A$ is also in contact with $B$.

Let us assess the correctness of this definition within one of the spatial ontologies encountered above.

For definiteness, we concentrate on $\tilde{R}$. (Recall that $\tilde{R}$ ambiguously denotes either $R$ or $\bar{R}$.)

**Observation 5.1.** Let $r_1, r_2 \in \tilde{R}$. Then $r_1 \leq r_2$ if and only if $\tilde{R}(C) \models \leq_C[r_1,r_2]$, where $\leq_C(x_1, x_2)$ is the $\mathcal{L}(C)$-formula $\forall z(C(x_1, z) \rightarrow C(x_2, z))$ .

**Proof:**
If $r_1 \leq r_2$ then $r_1^- \subseteq r_2^-$, so $t^- \cap r_1^- \neq \emptyset$ implies $t^- \cap r_2^- \neq \emptyset$ for any set $t$. Conversely, if $r_1 \not\leq r_2$, then $r_1 \cdot (-r_2)$ is non-empty, and we can certainly find a polygon $t$ lying in its interior, so that $t^- \cap r_1^- \neq \emptyset$, but $t^- \cap r_2^- = \emptyset$. □

Since the subset relation $\leq$ on $\tilde{R}$ is $\mathcal{L}(C)$-definable in $R(C)$, we may henceforth write the symbol $\leq$ in $\mathcal{L}(C)$-formulas, as a shorthand for $\leq_C$. It follows in addition that the Boolean functions $+$ , $\cdot$ and $-$, as well as the constants 0 and 1, are also $\mathcal{L}(C)$-definable, so again, we may henceforth write these symbols in $\mathcal{L}(C)$-formulas as a shorthand for their definitions. As a further shorthand, we write $x \oplus y = y$ for $x + y = z \wedge x.y = 0$.

Observation 5.1 tells us that Whitehead's definition of the part-whole relation is correct for the structures $\tilde{R}(C)$, in that the relevant formula really does pick out the subset relation as advertised. Again, note that this conclusion is a mathematical fact, not an intuitively plausible suggestion. We need no further assurance that the subset relation between elements of $\tilde{R}$ captures our pre-theoretic understanding of the part-whole relation, because we knew that anyway. This again illustrates the different ways of assessing definitions on the empiricist and rationalist approaches: while the former relies on appeal to pre-theoretic intuition, the latter substitutes rigorous argument connecting region-based accounts to familiar, point-based reconstructions of the defined relations.

Indeed, it is a virtue of our approach that it allows us to determine which spatial ontologies a proposed definition will work for. A little thought will show that observation 5.1 will work if we replace $\tilde{R}$ by $O$ or $Q$ (or their closed-plane versions); however, it will not work in the ontology $U$. In fact, we have:

**Observation 5.2.** There is no formula $\phi(x, y)$ in $\mathcal{L}(C)$ such that, for all $a, b \in U$, $U(C) \models \phi[a, b]$ if and only if $a \subseteq b$.

**Proof:**
Let $a$ be an open disc and $b$ be be its topological closure. Let $f : U \rightarrow U$ be the identity function except that $f(a) = b$ and $f(b) = a$. Then it is obvious that $f$ is a structure automorphism on $U(C)$, so that, for any formula $\phi(x, y)$, $U(C) \models \phi[a, b]$ if and only if $U(C) \models \phi[f(a), f(b)]$. The result then follows immediately. □

Moral: what is definable over one ontology may be indefinable over another.

Let us return to $\tilde{R}$. Having established that the part-whole relation is definable in $\tilde{R}(C)$, we turn now to the property of connectedness.

**Lemma 5.1.** *Let $r \in \tilde{R}$. Then $r$ is connected if and only if, for all nonempty, disjoint $r_1, r_2 \in \tilde{R}$ such that $r_1 + r_2 = r$, we have $\partial(r_1) \cap \partial(r_2) \cap r \neq \emptyset$.*

**Proof:**
Suppose $r$ is connected and $r_1$, $r_2 \in \tilde{R}$ satisfy $r_1 \oplus r_2 = r$. Let $q_i \in r_i$ $(i = 1, 2)$ and let $\alpha$ be a Jordan arc in $r$ from $q_1$ to $q_2$, which exists by the connectedness of $r$. Let $p$ be a point on $\alpha$ in $\partial(r_1)$. Then it is easy to see that $p \in \partial(r_1) \cap \partial(r_2) \cap r$. Conversely, suppose that $r$ is not connected. Let $r_1$ be any component of $r$, so that $r_1 \in \tilde{R}$. Let $r_2 = r \cdot -r_1$, so that $r_1 \oplus r_2 = r$. If $p \in \partial(r_1) \cap \partial(r_2)$, then, by lemma 5.2, we can find a point $q_i \in r_i$ and an end-cut $\alpha_i$ in $r_i$ from $q_i$ to $p$ $(i = 1, 2)$. If, in addition, $p \in r$, then $\alpha_1$ and $-\alpha_2$ form a Jordan arc in $r$ from $q_1$ to $q_2$, contradicting the fact that $q_1$ and $q_2$ lie in different components of $r$. Hence $\partial(r_1) \cap \partial(r_2) \cap r = \emptyset$. □

**Observation 5.3.** Let $r \in \tilde{R}$. Then $r$ is connected if and only if $R(C) \models c_C[r]$, where $c_C(x)$ is the $\mathcal{L}(C)$-formula

$$\forall x_1 \forall x_2 (x_1 \oplus x_2 = x \rightarrow \exists x'_1 \exists x'_2 (x'_1 \leq x_1 \wedge x'_2 \leq x_2 \wedge C(x'_1, x'_2) \wedge \neg C(x'_1 + x'_2, -x))).$$

**Proof:**
It is straightforward to show by lemma 5.2 that for nonempty and disjoint regions $r_1$ and $r_2$, $\partial(r_1) \cap \partial(r_2) \cap (r_1 + r_2) \neq \emptyset$ if and only if $r_1$ and $r_2$ satisfy the formula

$$\exists x'_1 \exists x'_2 (x'_1 \leq x_1 \wedge x'_2 \leq x_2 \wedge C(x'_1, x'_2) \wedge \neg C(x'_1 + x'_2, -(x_1 + x_2)))$$

in $\tilde{R}(C)$. The result then follows from lemma 5.1. □

Thus, we may henceforth write the symbol $c$ in $\mathcal{L}(C)$, understanding it as a shorthand for $c_C$. It follows of course, that, over polygons in both the open and closed planes, $\mathcal{L}(C)$ is at least as expressive as $\mathcal{L}(c, \leq)$.

## 5.2. Definability and non-definability in the language $\mathcal{L}(c, \leq)$

Consider how to define the property of being a Jordan region in the spatial ontology $\bar{R}(c, \leq)$. (Recall that a Jordan region is a region homeomorphic to an open disc.) To set up the definition, we need a specific fact about polygons and a general theorem.

**Definition 5.1.** Let $F$ be any set of points. An *end-cut* in $F$ from a point $p$ to a point $q$ is a Jordan arc from $p$ to $q$ lying entirely within $F$ except (possibly) for the final point $q$. A set $G$ is said to be *accessible* from $F$ if for every $p \in F$ and every $q \in G$, there is an end-cut in $F$ from $p$ to $q$.

**Lemma 5.2.** *If $s$ is a connected polygon, then $\mathcal{F}(s)$ is accessible from $s$.*

**Proof:**
Obvious. □

The following standard result can be found in Newman [15], ch. V, theorem 11.5.

**Theorem 5.1. (Converse of Jordan curve theorem)** *Let $G$ be a closed set which divides the closed plane into two residual domains, from each of which $G$ is accessible. Then $G$ is the locus of a Jordan curve.*

Now we can define the property of being a Jordan region in the closed plane.

**Observation 5.4.** Let $j(x)$ be the formula

$$j(x) := c(x) \wedge x \neq 0 \wedge c(-x) \wedge x \neq 1.$$

Then, for all $s \in \bar{R}$, $\bar{R}(c, \leq) \models \phi[s]$ if and only if $s$ is a Jordan region.

**Proof:**
If $s \in \bar{R}$ is a Jordan region, then it is obvious that $\bar{R}(c, \leq) \models j[s]$. Conversely, suppose, $\bar{R}(c, \leq) \models j[s]$. Then $\mathcal{F}(s)$ divides the closed plane into two residual domains which are connected polygons. By lemma 5.2, $\mathcal{F}(s)$ is accessible from each of these residual domains, and so, by theorem 5.1, $s$ is a Jordan region. □

Again, it is important to appreciate the care we took in checking that $j(x)$ really does define the property of being a Jordan region, because the definition does not work if we liberalize our collection of regions. Consider, for example the spatial ontology $\bar{O}(c, \leq)$ of all regular open subsets of the closed plane. Then the regular open set $a_2$ in fig. 3 satisfies $j(x)$ in $\bar{O}(c, \leq)$, but is not a Jordan region. The restriction to polygons—or, more precisely, the restriction to regions for which lemma 5.2 holds—is thus essential.

Observation 5.4 is further qualified to apply only to the *closed* plane. It is instructive to see why this qualification is essential also. Any two Jordan regions are similarly situated in the closed plane: if $s_1$ and $s_2$ are such regions, there is a homeomorphism of the closed plane onto itself taking $s_1$ to $s_2$. It then follows that $j(x)$ is a *complete* formula in $\mathrm{Th}(\bar{R}(c, \leq))$, in the following sense:

**Definition 5.2.** Let $T$ be a theory. We say that a formula $\phi$ is *complete* in $T$ if, for all formulas $\theta$, exactly one of $\phi \rightarrow \theta \in T$ and $\phi \rightarrow \neg\theta \in T$ hold.

**Observation 5.5.** The formula $j(x)$ defined above is complete in $\mathrm{Th}(\bar{R}(c, \leq))$.

**Proof:**
It is easy to show that lemma 4.7 continues to hold when $R$ is replaced by $\bar{R}$. The result is then immediate. □

Now since $R(c, \leq)$ and $\bar{R}(c, \leq)$ are isomorphic, their theories are the same, so any formula complete in $\mathrm{Th}(\bar{R}(c, \leq))$ will be complete in $\mathrm{Th}(R(c, \leq))$, including the formula $j(x)$. This establishes the following non-definability result.

**Observation 5.6.** There is no formula $\phi(x) \in \mathcal{L}(c, \leq)$ such that, for all $r \in R$, $r$ is a Jordan region if and only if $\mathbb{R}(c, \leq) \models \phi[r]$.

**Proof:**
Let $r$ be any Jordan region in the open plane. Visibly, both $r$ and $-r$ satisfy $j(x)$, which is complete in $\mathrm{Th}(R(c, \leq))$, so $r$ and $-r$ satisfy all the same formulas in $R(c, \leq)$. But $r$ is a Jordan region and $-r$ is not. □

It easily follows by the same argument that the property of being bounded is also not definable in $R(c, \leq)$.

Thus, definability results depend not only on the details of which sets of points we allow as regions, but also on the global topological properties of the space we locate them in. We note in passing that the notion of an atomic formula will play an important role in the sequel.

To round off this section, we show that the contact relation is definable in the structure $\bar{R}(c, \leq)$. Hence, for polygons in the closed plane, the language $\mathcal{L}(c, \leq)$ is at least as expressive as the language $\mathcal{L}(C)$. This observation is instructive, because it shows how our approach to developing region-based theories of space sometimes uses highly non-trivial mathematical results.

**Theorem 5.2. (Newman [15], ch. V, theorem 9.2)** *Let $F_1$ and $F_2$ be closed sets in the closed plane. If $F_1 \cap F_2$ is connected, then two points which are connected in the complement of $F_1$ and also in the complement of $F_2$ are connected in the complement of $F_1 \cup F_2$.*

**Lemma 5.3.** *Let $s_1, s_2, t \in \bar{R}$ with $-(s_1 + t)$, $-(s_2 + t)$ and $t$ all connected, and $s_1^- \cap s_2^- = \emptyset$ . Then $-(s_1 + s_2 + t)$ is also connected.*

**Proof:**
Set $F_i = (s_i + t)^-$ $(i = 1, 2)$. Thus, the complement of $F_i$ is $-(s_i + t)$ $(i = 1, 2)$, and the complement of $F_1 \cup F_2$ is $-(s_1 + s_2 + t)$. Moreover, since $t$ is connected, so is $t^-$, whence $F_1 \cap F_2 = (s_1 + t)^- \cap (s_2 + t)^- = (s_1^- \cup t^-) \cap (s_2^- \cup t^-) = (s_1^- \cap s_2^-) \cup t^- = t^-$ is connected. Choose any distinct $p, q \in -(s_1 + s_2 + t)$, and the result follows by theorem 5.2. □

In the sequel we write $\pi(y_1, y_2)$ to abbreviate the formula

$$\exists z(c(-(y_1 + z)) \wedge c(-(y_2 + z)) \wedge c(z) \wedge \neg c(-(y_1 + y_2 + z))).$$

**Observation 5.7.** Let $r_1, r_2 \in \bar{R}$. Then $r_1^- \cap r_2^- \neq \emptyset$ if and only if $\bar{R}(c, \leq) \models C_{c,\leq}[r_1,r_2]$, where $C_{c,\leq}(x_1, x_2)$ is the formula:

$$\exists y_1 \exists y_2 (y_1 \leq x_1 \wedge y_2 \leq x_2 \wedge \pi(y_1, y_2)).$$

**Proof:**
The if-direction is immediate given lemma 5.3. For the only-if direction, let us say that $s_1$ and $s_2$ *form a bow tie* if they are similarly situated to the arrangement shown in figure 7. Visibly, $\bar{R}(c, \leq) \models \pi[s_1, s_2]$. Moreover, it is easy to see by lemma 5.2 that if $r_1^- \cap r_2^- \neq \emptyset$ then we can find $s_1 \leq r_1$, $s_2 \leq r_2$ such that $s_1$ and $s_2$ form a bow tie, whence $\bar{R}(c, \leq) \models \phi_C[r_1, r_2]$. □

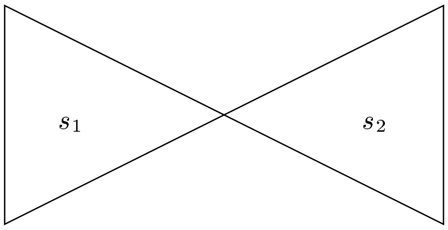



Figure 7. Regions $s_1$ and $s_2$ forming a bow tie.

Thus, the contact relation is definable in the spatial ontology $\bar{R}(c, \leq)$. In this case, the defining formula is so complicated that informal judgement is not capable of determining whether it really works as advertised. To do this, we rely on the nontrivial and nonobvious fact about the topology of the closed plane reported in theorem 5.2. We remark that the restriction to the closed plane is essential: the relation $C(x, y)$ is in fact not definable in the spatial ontology $R(c, \leq)$.

## 5.3. Expressive completeness

We know from lemma 4.7 that similarly situated tuples satisfy the same formulas in $R(c, \leq)$, a result which clearly generalizes to $X(\sigma)$ where $X$ is any topologically homogeneous set of regions (in the sense of lemma 4.5) and $\sigma$ is any set of topological primitives. We also know from the above discussion that the converse does not hold: a Jordan region in $R$ and its pseudo-complement are not similarly situated, yet they both satisfy the formula $j(x)$ and hence satisfy all the same formulas of $\mathcal{L}(c, \leq)$. The question naturally arises as to whether this converse result nevertheless holds for the other topological signatures and spaces considered here. The answer was first provided, in a restricted form, by Papadimitriou, Suciu and Vianu [16], and later, using a very different line of argument, by Pratt and Schoop [18]. We present the results reported in the latter source here.

**Definition 5.3.** Let $A$ be a set of subsets of some topological space. A formula in $\mathcal{L}(\sigma)$ is said to be *topologically complete* in the the spatial ontology $A(\sigma)$ if, for all $n$-tuples $\bar{a}$ and $\bar{b}$ in $A$, $A(\sigma) \models \phi[\bar{a}]$ and $A(\sigma) \models \phi[\bar{b}]$ implies $\bar{a} \sim \bar{b}$ .

**Observation 5.8.** Every tuple in $R$ satisfies a topologically complete formula in $R(C)$. Every tuple in $\bar{R}$ satisfies a topologically complete formula in $\bar{R}(C)$. Every tuple in $\bar{R}$ satisfies a topologically complete formula in $\bar{R}(c, \leq)$.

The proofs are too involved to be reproduced here.

Observation 5.8 can be seen as a kind of expressiveness result. It says that, in the open plane, any tuple is topologically completely characterized by a single $\mathcal{L}(C)$-formula; and, in the closed plane, any tuple is topologically completely characterized by a single $\mathcal{L}(C)$-formula and also by a single $\mathcal{L}(c, \leq)$-formula. Observation 5.5 tells us that, by contrast, it is possible to have non-similarly-situated tuples in the open plane which are not separated by any $\mathcal{L}(c, \leq)$-formula.

In the nontopological language $\mathcal{L}(\text{conv}, \leq)$, a similar result obtains. The following result is taken from Pratt [19].

**Definition 5.4.** Let $A$ be a set of subsets of $\mathbb{R}^m$. A set of formulas $\Phi(\bar{x})$ in $\mathcal{L}(\sigma)$ is said to be *affine complete* in the the spatial ontology $A(\sigma)$ if, for all $n$-tuples $\bar{a}$ and $\bar{b}$ in $A$, $A(\sigma) \models \Phi[\bar{a}]$ and $A(\sigma) \models \Phi[\bar{b}]$ implies that there is an affine transformation of $\mathbb{R}^m$ mapping $\bar{a}$ to $\bar{b}$. A formula $\phi(\bar{x})$ is said to be *affine complete* if the singleton $\{\phi(\bar{x})\}$ is affine complete.

**Observation 5.9.** Every tuple in $Q$ satisfies an affine complete formula in $Q(\text{conv}, \leq)$. Every tuple in $R$ satisfies an affine complete set of formulas in $R(\text{conv}, \leq)$.

Again, The proofs are too involved to be reproduced here.

Observation 5.9 says that, in the rational polygonal ontology, any tuple is affine-characterized by a single $\mathcal{L}(\text{conv}, \leq)$-formula; moreover, in the real polygonal ontology, any two non-affine-equivalent tuples are separated by a $\mathcal{L}(\text{conv}, \leq)$-formula. We note that Davis, Gotts and Cohn [8] present a similar result for a language with an additional primitive but more restrictive logical resources.

## 6. Alternative ontologies

It is a custom in our society to amuse children by telling them stories in which the protagonist is given his heart's desire, but—as it quickly transpires—in a way that leads to his downfall and destruction. By the time he realizes it, of course, it is too late. Region-based theories of space are also in part motivated by a wish—the wish for a region-based spatial ontology that is a real alternative to the familiar point-based ontology. *We* are now in a good position to see just how this wish might be fulfilled. In particular, we understand the importance of characterizing region-based ontologies precisely, because different ways of building regions in terms of points lead to different spatial ontologies with different logical properties. Having made this discovery, we can return to the quest for alternative, region-based ontologies with eyes more open to what is involved.

For definiteness, consider the spatial ontology $Q(c, \leq)$. We understand the search for an alternative spatial ontology as follows: find a structure $\mathfrak{A}$ such that $\mathfrak{A} \equiv Q(c, \leq)$ but $\mathfrak{A} \not\simeq Q(c, \leq)$. Such a structure may be *truly* region-based, in the sense that its only contact with the familiar point-based model of space is that it makes the same sentences true as $Q(c, \leq)$. And who knows what interesting alternative spatial ontologies await discovery? In the sequel, we shall indeed find what we are seeking, but when we do so, we shall realize that we never really wanted it in the first place.

First, a very easy observation. The upward Löwenheim-Skolem-Tarski theorem implies that any infinite structure has elementarily equivalent structures of all infinite cardinalities. Since structures of different cardinalities cannot be isomorphic, $Q(c, \leq)$ certainly has non-isomorphic elementarily equivalent structures. Indeed, we have already found one: the model $R(c, \leq)$. But $R(c, \leq)$ is hardly a very interesting alternative spatial ontology to $Q(c, \leq)$, because $Q(c, \leq)$ is a (proper) elementary submodel of $R(c, \leq)$. That is: $R(c, \leq)$ properly contains $Q(c, \leq)$, with the extra elements making no difference to the $\mathcal{L}$-expressible properties satisfied by tuples of elements from $Q$. So $R(c, \leq)$ is an alternative spatial ontology, but not really what we were looking for: it is just $Q(c, \leq)$ with a lot of irrelevant junk thrown in.

But we can do better than this. We begin with a particular observation on the structure $Q(c, \leq)$.

**Definition 6.1.** If $\bar{x}$ is a tuple of variables, then a *type* in $\bar{x}$ is a complete set of formulas with free variables $\bar{x}$—that is a set $\Phi$ such that, for every formula $\phi$ with free variables $\bar{x}$, either $\phi \in \Phi$ or $\neg\phi \in \Phi$. A theory $T$ is said to *have* a type $\Phi$ if $\Phi$ is consistent with $T$.

**Lemma 6.1.** $\mathrm{Th}(Q(c, \leq))$ *has (at least) countably many types in the single variable* $x$.

**Proof:**
It is easy to see that, for every positive integer $m$, the formula $\phi_m(x)$

$$\exists z_1 \ldots \exists z_m \left( \bigwedge_{1 \leq i \leq m} (c(z_i) \wedge z_i \neq 0) \wedge \bigwedge_{1 \leq i < j \leq m} \neg c(z_i + z_j) \wedge x = \sum_{1 \leq i \leq m} z_i \right)$$

is satisfied by all and only those elements in $Q$ having exactly $m$ components. Hence, the $\phi_m(x)$ are all satisfied in $Q(c, \leq)$; so each can be extended to a type $\Gamma_m(x)$ of $\mathrm{Th}(Q(c, \leq)$. But the $\phi_m(x)$ are also pairwise mutually exclusive in $\mathrm{Th}(Q(c, \leq))$; so no two of them can be extended to the same type in $\mathrm{Th}(Q(c, \leq))$. Hence, $\mathrm{Th}(Q(c, \leq))$ has countably many types in $x$. □

Now, we deploy a standard result from model theory to establish that there are indeed countable alternative spatial ontologies to $Q(c, \leq)$. (In this section, standard theorems are from Chang and Keisler [4].)

**Definition 6.2.** A theory $T$ is said to be $\omega$*-categorical* if all countable models of $T$ are isomorphic.

**Theorem 6.1. (Chang and Keisler: 2.3.13)** *Let $T$ be a complete theory. Then $T$ is $\omega$-categorical if and only if, for each $n$, $T$ has only finitely many types in $x_1, \ldots, x_n$.*

Whence:

**Observation 6.1.** The theory $\mathrm{Th}(Q(c, \leq)) = \mathrm{Th}(\tilde{R}(c, \leq))$ is not $\omega$-categorical.

Having established that there *are* alternative spatial ontologies—even countable ones, the slightly more tricky part is to show that none of them is any good. To see what is going on here, we need some more ideas from model theory. We have already encountered the concept of a *complete* formula with respect to a theory $T$: it is one which, given $T$, fixes all the expressible properties of any tuple which satisfies it. We have also encountered an example of a complete formula with respect to $\mathrm{Th}(Q(c, \leq)$: the formula $j(x)$.

**Definition 6.3.** A model $\mathfrak{A}$ is said to be *atomic* if any $n$-tuple $\bar{a}$ in $A$ satisfies a formula $\phi(\bar{x})$ in $\mathfrak{A}$ such that $\phi(\bar{x})$ is complete in $\mathrm{Th}(\mathfrak{A})$.

We have the following observations regarding the structure $Q(c, \leq)$:

**Observation 6.2.** $Q(c, \leq)$ is an atomic model.

A proof of this result (not difficult) is given in Pratt and Lemon [20]. Next, we again deploy a standard result from model theory.

**Definition 6.4.** A structure $\mathfrak{A}$ is said to be *elementarily embeddable* in a structure $\mathfrak{B}$ if $\mathfrak{A}$ is isomorphic to an elementary submodel of $\mathfrak{B}$. A model $\mathfrak{A}$ is said to be *prime* if, for any model $\mathfrak{B}$, $\mathfrak{A} \equiv \mathfrak{B}$ implies that $\mathfrak{A}$ is elementarily embeddable in $\mathfrak{B}$.

**Theorem 6.2. (Chang and Keisler: 2.3.4)** *A model is countable atomic if and only if it is prime.*

It easily follows:

**Observation 6.3.** The structure $Q(c, \leq)$ is elementarily embeddable in any structure $\mathfrak{A}$ elementarily equivalent to $Q(c, \leq)$.

So, the alternative spatial ontologies are all elementary extensions of copies of $Q(c, \leq)$! That is: any such ontology $\mathfrak{A}$ contains a copy of $Q(c, \leq)$, with the extra elements making no difference to any expressible properties and relations over the elements of that copy. It is in this sense that the alternative spatial ontologies are no good.

Another question that arises is whether any genuine alternatives to $Q(c, \leq)$ also have this property: after all, corollary 6.3 may have shown that the familiar ontology $Q(c, \leq)$ is (in some sense) a maximally simple spatial ontology; but maybe there are other maximally simple ontologies too. In fact, however, even this is false, as the following standard result from model theory shows.

**Theorem 6.3. (Chang and Keisler: 2.3.3)** *If $\mathfrak{A}$ and $\mathfrak{B}$ are countable atomic models and $\mathfrak{A} \equiv \mathfrak{B}$, then $\mathfrak{A} \simeq \mathfrak{B}$.*

Thus, $Q(c, \leq)$ is not *a* maximally simple spatial ontology for its theory, it is, up to isomorphism, *the* maximally simple spatial ontology for its theory.

We end this section with a warning. The result we have proved about $Q(c, \leq)$ does not straightforwardly generalize to other ontologies! In particular, as far as the author is aware, nothing comparable is known about the possible models of $O(c, \leq)$ or $U(c, \leq)$.

## 7. Conclusion

This paper has contrasted what we called the empiricist and rationalist approaches to developing region-based theories of space. The empiricist approach, which informs most early work on the subject, can be characterized as follows. First, select a group of primitive spatial relations corresponding to familiar spatial concepts, and illustrate their meaning with some examples. Next, write down axioms to govern these primitives. Finally, propose various definitions for a range of familiar spatial relations not included among the chosen set of primitives. The appropriateness of these axioms and definitions is to be judged on the basis of their ability to chime with pre-theoretic intuition and, perhaps, of their success in a larger theory of commonsense spatial reasoning. In particular, no attempt is made to derive or deduce the axioms of the theory, because there is nothing to derive or deduce them from.

The rationalist approach, by contrast, can be characterized as follows. First, select a group $\sigma$ of predicate letters to represent primitive spatial relations corresponding to familiar spatial concepts. Next, using some familiar, point-based model of space, select a set $A$ of subsets of that space to count as the *regions* recognized by the theory. Finally, interpret the predicate letters in $\sigma$ over the regions $A$ using the standard definitions so as to obtain a structure $A(\sigma)$. Having obtained this structure, we can then systematically investigate its properties as a structure. In particular, we can investigate the extent to which the set $A$ of regions can be varied with out changing $\mathrm{Th}(A(\sigma))$; we can try to characterize $\mathrm{Th}(A(\sigma))$ axiomatically; we can determine whether various relations over $A$ can be defined by formulas of $\mathcal{L}(\sigma)$; we can provide criteria for when tuples of $A$ satisfy the same formulas of $\mathcal{L}(\sigma)$; and we can look for alternative models of $\mathrm{Th}(A(\sigma))$ and determine their relationships to $A(\sigma)$.

The results reported in this paper illustrate some of these questions and the mathematical techniques required to answer them. It thus constitutes an extended advertisement for the rationalist approach, and an attack on its empiricist competitor. Most importantly, we have seen how the rationalist approach constrains the selection of axioms and the definition of non-primitive spatial relations by relating region-based theories of space to familiar point-based models whose usefulness as representations of the space we inhabit is beyond doubt. And we have seen how the rationalist approach gives us surprising insights into what region-based ontologies are possible, and how they relate to the point-based models with which we are familiar. To be sure, the empiricist approach enjoys the virtues of simplicity and freedom. Perhaps for that reason, its representatives have historically been the prime movers in the development of region-based theories of space. But those very virtues, we claim, limit what that approach can ultimately achieve. If the past development of region-based theories of space has been dominated by the empiricists, then its future belongs to the rationalists.

## References

[1] N. Asher and L. Vieu. Toward a geometry of common sense – A semantics and a complete axiomatization of mereotopology In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI 95)*, pages 846–852, 1995.

[2] L. Biacino and G. Gerla. Connection structures. *Notre Dame Journal of Formal Logic*, 32(2):242–247, 1991.

[3] S. Borgo, N. Guarino, and C. Masolo. A pointless theory of space based on strong connection and congruence. In L. C. Aiello, J. Doyle, and S. C. Shapiro, editors, *Principles of Knowl-*

*edge Representation and Reasoning: Proceedings of the Fifth International Conference (KR 96)*, pages 220–229, San Francisco, CA., 1996. Morgan Kaufmann Publishers.

[4] C. C. Chang and H. J. Keisler. *Model Theory.* North Holland, Amsterdam, 3rd edition, 1990.

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

[6] B. L. Clarke. A calculus of individuals based on "connection". *Notre Dame Journal of Formal Logic*, 22(3):204–218, 1981.

[7] A. G. Cohn, L. K. Schubert, and S. C. Schubert, editors. *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, San Francisco, CA, 1998. Morgan Kaufman.

[8] Ernest Davis, Nicholas Mark Gotts, and Anthony G. Cohn. Constraint networks of topological relations and convexity. *Constraints*, 4(3):241–280, 1999.

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

[10] Christoph Dornheim. Undecidability of plane polygonal mereotopology. In Cohn et al. [7], pages 342–353.

[11] N. M. Gotts, J. M. Gooday, and A. G. Cohn. A connection based approach to commonsense topological description and reasoning. *Monist*, 79(1):51–75, 1996.

[12] L. Henkin, P. Suppes, and A. Tarski, editors. *The Axiomatic Method, with Special Reference to Geometry and Physics.* North-Holland Publishing Co., Amsterdam, 1959.

[13] P. T. Johnstone. *Stone Spaces.* Cambridge University Press, 1982.

[14] O. Lemon. Semantical foundations of spatial logics. In L. C. Aiello, J. Doyle, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, pages 212–219, San Francisco, CA., 1996. Morgan Kaufmann Publishers.

[15] M. H. A. Newman. *Elements of the Topology of Plane Sets of Points.* Cambridge University Press, Cambridge, 1964.

[16] C. H. Papadimitriou, D. Suciu, and V. Vianu. Topological queries in spatial databases. In *Proceedings of PODS'96*, pages 81–92. ACM, 1996.

[17] I. Pratt and D. Schoop. A complete axiom system for polygonal mereotopology of the real plane. *Journal of Philosophical Logic*, 27(6):621–658, 1998.

[18] I. Pratt and D. Schoop. Expressivity in polygonal, plane mereotopology. *Journal of Symbolic Logic*, 65(2):822–838, 2000.

[19] Ian Pratt. First-order spatial representation languages with convexity. *Spatial Cognition and Computation*, 1:181–204, 1999.

[20] Ian Pratt and Oliver Lemon. Ontologies for plane, polygonal mereotopology. *Notre Dame Journal of Formal Logic*, 38(2):225–245, 1997.

[21] D. A. Randell, Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. In B. Nebel, C. Rich, and W. Swartout, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR '92)*, pages 165–176, Los Altos, CA, 1992. Morgan Kaufmann Publishers.

[22] Jochen Renz. A canonical model of the region connection calculus. In Cohn et al. [7], pages 330–341.

[23] Alfred Tarski. What is Elementary Geometry? In L. Henkin, P. Suppes, and A. Tarski, editors, *The Axiomatic Method, with Special Reference to Geometry and Physics*, pages 16–29. North-Holland Publishing Co., Amsterdam, 1959.

[24] A. N. Whitehead. *Process and Reality.* The MacMillan Company, New York, 1929.