

Defining Relations: A General Incremental Approach with Spatial and Temporal Case Studies

Brandon BENNETT ^{a,1}, Heshan DU ^a, Lucía GÓMEZ ÁLVAREZ ^a,
Anthony G. COHN ^{a,2}

^a *School of Computing, University of Leeds, LS9 2JT, UK*

Abstract. This paper aims to lay a foundation for a systematic study of mechanisms for construction of definitions within a formal theory, by investigating operators for incremental construction of definitions of new relations from an existing set of primitives and previously defined relations. To illustrate our method, we apply it to two of the best known relation sets studied in KRR: Allen's Interval Algebra and Region Connection Calculus. We also show that systematic exploration of definitional possibilities can yield interesting insights into relation sets that were originally defined in a more *ad hoc* way, and opens the possibility for discovering new vocabulary for extending or refining existing calculi or for developing completely new calculi.

Keywords. Relations, Definitions

1. Introduction

The literature on Knowledge Representation (KR) since the mid-80s shows increasing interest in specification and investigation of particular formal vocabulary sets designed to encode and support reasoning about certain limited but significant kinds of information. In many cases the formal vocabulary considered is a set of relations and the focus has been on reasoning with sets of relational constraints [1,2,3,4]. Among the huge variety of relation sets, the 13 temporal interval relations of Allen [5] and the topological relations of Region Connection Calculus (RCC) [1] are typical examples that are widely known in the field of KR and have also been extensively employed in ontology construction.

A pervasive current in information sciences is a growing awareness of the importance of *definitions* in developing computational infra-structure for representing, interpreting and combining information. This can be seen in the increasing use of ontologies to specify conceptual vocabularies that characterise the meaning of data and make explicit the information it conveys. A large number of theories have also been proposed within which a vocabulary of concepts has been defined in terms of a small set of primitives [5,1,6,7]. Nevertheless there seems to be little work focusing directly on the ways in which definitions can be constructed.

¹Corresponding Author: B.Bennett@leeds.ac.uk

²Support of EU FP7 project 600623 is gratefully acknowledged.

The main novelty of this paper is to lay a foundation for systematic study of mechanisms for construction of definitions within a formal theory. We apply this to some of the best known relation sets used in KR. This will illustrate our method of generating and analysing definitions. We also show that systematic exploration of definitional possibilities can yield interesting insights into relation sets that were originally defined in a more *ad hoc* way.

Historically there has been considerable interest in whether a given set of primitives is sufficient to define the concepts of a theory. This has been particularly so in the field of geometry [8,9,10,11]. More general methods have been developed for establishing definability with respect to arbitrary theories [12] formulated in classical logic. This work has primarily concentrated on the semantic question of whether a given predicate is definable from a given set of primitives, without concern for the particular mechanisms by which definitions are constructed.

By contrast, this paper is concerned with the iterative syntactic process of constructing definitions from primitive relations. We shall investigate methods for incremental construction of definitions *via* a limited set of operations. This framework will allow us to address questions such as: ‘What is the ‘shortest’ way to reach a given definition?’; ‘How does the background theory of its primitives affect the space of possible relations that can be defined?’; and, ‘Is there a limit to the number of relations that can be defined from a given primitive relation?’

We shall be concerned mainly with representational issues, rather than reasoning. However, this kind of definitional analysis may yield insights that can be applied to automated reasoning. One simple way of applying our approach is to take the axioms and definitions of an existing KR formalism and test whether any of the definitions of relations used the theory could be replaced by simpler (i.e. lower level) definitions. For example, our approach has enabled us to discover definitions of certain key relations in RCC [1], that are much simpler but logically equivalent to previously published definitions.

In the next section we consider ways that possible definitions can be generated systematically from given primitives. We then consider some sets of all possible relations that can be defined by a particular definitional framework, using a limited number of definitional operations. In sections 4 and 5 we turn our attention to the particular case of of the Allen interval calculus and RCC.

2. Ways to Generate Relation Sets

There are a number of ways in which one might generate sets of relations, including:

1. Intuitively identify significant cases from a space of possibilities and then invent formal definitions which seem to capture these intuitions.
2. Consider Boolean combinations of certain conditions; and delete cases that are impossible according to some constraints.
3. Systematically consider all possible ways that relations can be defined from basic primitives.

In case (1), although the definitions constructed are formal, the process by which they are identified is largely *ad hoc*. The development of the RCC calculi [1,13] followed this style of approach. Often, as in the case of RCC, relations will be defined from a small set

of primitives (e.g. the RCC *connection* relation), but this need not be the case. An example of (2) is the 9-intersection method of [14], where a set of topological relations is identified by considering possible combinations of intersection conditions that can hold between the exteriors, interiors and boundaries of two regions. We are not aware of previous work that has employed approach (3) as a means to discover and establish definitions for sets of relations intended for use in a knowledge representation language or ontology. The focus of this paper is to establish a basis from which this approach can be developed and explore some directions it can be taken. In particular we shall consider how a space of definable relations can be generated from primitive relations within the context of some background theory.

2.1. Binary Incremental Relation Definition Generator

A simple idea for incremental construction of definitions of relations is to start from a set of primitive relations and consider operations that construct a new relation from a combination of any *two* (previously defined or primitive) relations.

Let us start with a set \mathcal{P} of binary primitive relations. For any relation $R \in \mathcal{P}$, we consider two *unary* operators: negation and *converse*. $\neg R$ is the negation of R and R^\smile refers to the *converse* of R — i.e. $R^\smile xy$ is equivalent to Ryx . We say that $\neg R$, R^\smile and $\neg R^\smile$ are *cognates* of R . When writing first order formulae we usually avoid using the \smile symbol by simply transposing the arguments to which the relation is applied. Since the polarity and argument ordering of a predicate do not affect what can be defined from it, we assume the unary negation and converse operations do not increase the definitional complexity (level) of a relation. We stipulate that all relations in \mathcal{P} and all cognates of these relations are *definable at level 0*. More generally, we specify the *level* of a defined relation as follows:

- For every $R \in \mathcal{P}$, $R(x, y)$ is a level 0 relation.
- If a relation $R(x, y)$ is definable at level n , then its cognates $\neg R(x, y)$, $R(y, x)$ and $\neg R(y, x)$ are also definable at level n .
- For any two relations R and S that are definable at levels m and n respectively, the following relations are definable at Level $l = \text{Max}(m, n) + 1$ by the following binary relation combining operations:

$$\text{(Conj)} \quad (R(x, y) \wedge S(x, y)) \quad \text{(Compos)} \quad \exists z[R(x, z) \wedge S(z, y)]$$

Because we do not count application of unary operations as definitional steps, various other binary combinations count as a single step; e.g. $(R(x, y) \vee S(x, y)) \equiv \neg(\neg R(x, y) \wedge \neg S(x, y))$ and $\forall z[R(x, z) \rightarrow S(z, y)] \equiv \neg\exists z[R(x, z) \wedge \neg S(z, y)]$. A relation can be definable at different levels. We say R is a Level m relation, if for any $n < m$, R is *not* definable at level n .

For example, if **Parent** is a primitive relation, then **Grandparent**(x, y) $\equiv_{\text{def}} \exists z[\text{Parent}(x, z) \wedge \text{Parent}(z, y)]$ can be defined as a Level 1 relation.

These operations specify a set of possible relations that can be defined from \mathcal{P} , whose definitions are syntactically distinct. Of course there can be distinct members of this set that are semantically equivalent either in a strong sense of being logically equivalent (hence, true in exactly the same models) or in a weaker sense of being true in all models satisfying some background axioms (e.g. general properties of the base relations or constraints on the elements of the domain).

Our characterisation of relational definitions by sequential application of operators is closely related to the *Relation Algebra* (RA) formalism, which considers relations as elements of a Boolean algebra augmented with *composition* and *converse* operators. Such an algebra is constrained to obey axioms first specified by [15] and later investigated in great detail in [16]. RAs have been found to be valuable in the analysis of composition based reasoning algorithms for temporal relations [17] and have been applied to describing sets of spatial relations [18]. Although RAs provide an elegant formalism for describing combinations of relations, we shall use classical first order logic as our main representation language. This is because, at least in our initial investigations, we are using standard first-order automated reasoning to identify equivalences among generated relations.

2.2. More Expressive Relation Generation Mechanisms

As detailed in [16], the language of equations between relation algebraic terms can be regarded as a strictly less expressive sub-language of first-order logic. It follows from this that it is possible to define relations in first-order logic that are not definable using only the operators we have specified. Nevertheless, [16] note that in the presence of certain domain constraints, such as a minimum size of the domain, a first-order formula that is inexpressible in unconstrained Relation Algebra may become equivalent to some relation algebraic formula. More generally, within the context of a particular background theory it may be that all first-order definable relations can be defined using the operations specified above. Indeed, this is known to be true for Allen's interval relations [5].

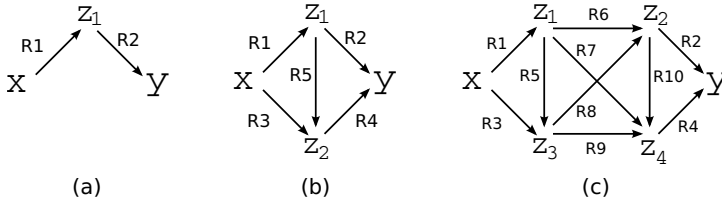


Figure 1. Defining new relations from networks of previously defined relations.

However, depending on the theory and kinds of definition we are exploring, we may wish to generalise the definition constructions that are allowed. Figure 1 illustrates how new relations can be constructed by *composition* (a), *conjoined composition* (b) and *network composition*. In each case the variables z_i are existentially quantified. Conjoined composition allows one to constrain a conjunction of two composition operations by specifying a relation (R5) between the two existentially quantified variables. Case (c) is very general since the internal constraints R6–R9 could themselves be complex defined relations. We conjecture that network composition gives the full expressivity of first-order logic.

3. Exhaustive Incrementally Generated Relation Sets

3.1. Freely Generated Level 1 Relations from R

Let us consider the generation of relations from an arbitrary primitive relation R by means of the binary definition forming operations. After applying one binary relation combining

Table 1. Level 1 relation definitions generated by negation, conjunction and composition.

N	Relation	Conv	Symmetry	Allen
			e.g. R is C	R is M
0	\perp	(0)	\perp	\perp
1	$R(x, y) \wedge R(y, x)$	(1)	$C(x, y)$	\perp
2	$\neg R(x, y) \wedge R(y, x)$	3	\perp	$Mi(x, y)$
3	$R(x, y) \wedge \neg R(y, x)$	2	\perp	$M(x, y)$
4	$\neg R(x, y) \wedge \neg R(y, x)$	(4)	$\neg C(x, y)$	$\neg MMi(x, y)$
5	$\exists z[R(x, z) \wedge R(y, z)]$	(5)	$IndC(x, y)$	$FS(x, y)$
6	$\exists z[\neg R(x, z) \wedge R(y, z)]$	7	$\neg Pi(x, y)$	$\neg FS(x, y)$
7	$\exists z[R(x, z) \wedge \neg R(y, z)]$	6	$\neg P(x, y)$	$\neg FS(x, y)$
8	$\exists z[\neg R(x, z) \wedge \neg R(y, z)]$	(8)	$\neg JU(x, y)$	\top
9	$\exists z[R(z, x) \wedge R(y, z)]$	13	$IndC(x, y)$	$A(x, y)$
10	$\exists z[\neg R(z, x) \wedge R(y, z)]$	15	$\neg Pi(x, y)$	\top
11	$\exists z[R(z, x) \wedge \neg R(y, z)]$	14	$\neg P(x, y)$	\top
12	$\exists z[\neg R(z, x) \wedge \neg R(y, z)]$	16	$\neg JU(x, y)$	\top
13	$\exists z[R(x, z) \wedge R(z, y)]$	9	$IndC(x, y)$	$B(x, y)$
14	$\exists z[\neg R(x, z) \wedge R(z, y)]$	11	$\neg Pi(x, y)$	\top
15	$\exists z[R(x, z) \wedge \neg R(z, y)]$	10	$\neg P(x, y)$	\top
16	$\exists z[\neg R(x, z) \wedge \neg R(z, y)]$	12	$\neg JU(x, y)$	\top
17	$\exists z[R(z, x) \wedge R(z, y)]$	(17)	$IndC(x, y)$	$SS(x, y)$
18	$\exists z[\neg R(z, x) \wedge R(z, y)]$	19	$\neg Pi(x, y)$	$\neg SS(x, y)$
19	$\exists z[R(z, x) \wedge \neg R(z, y)]$	18	$\neg P(x, y)$	$\neg SS(x, y)$
20	$\exists z[\neg R(z, x) \wedge \neg R(z, y)]$	(20)	$\neg JU(x, y)$	\top

operation, together with any number of the unary negation and converse operators, we get 42 *distinct* Level 1 relations, including \perp (the empty relation) and \top (the universal relation). Specifically, we get the 21 relations given in Table 1 (Column 2) and also the negations of these.

Among the 21 Level 1 relations, 7 relations are syntactically symmetric (the converse of a symmetric relation is itself) and the remaining 14 consist of 7 non-symmetric relations and their converses. When constructing Level 2 definitions, we need only consider combinations drawn from a set of 13 Level 1 relations (6 symmetric relations and 7 non-symmetric relations) plus the primitive R . It is easy to check that relations \perp and \top are not useful since however they are combined with another relation, R , the result is always equivalent to either \perp , \top or some cognate of R . From the 14 non-symmetric relations we only need consider one of each pair.

We have also analysed all ways in which two distinct relations R and S can be combined by our binary relation generation operators. We do not give details here, but it turns out that 96 distinct Level 1 relations can be constructed involving two different relations. A total of 178 Level 1 relations can be constructed using either just R just S or both R and S .³

³We have 96 relations constructed using both R and S and 42 with each of R of S . $96 + 2 * 42 = 180$; but \top and \perp are each generated both from R and S , so we need to subtract 2.

3.2. Constraints on Relations and the Domain

We now consider how to generate such specific sets in a systematic way. A systematic generation has advantages over an ad-hoc process, such as ensuring that every definition is at its lowest possible level and providing insight into the way a relation set builds up. The generation may be done by applying constraints on the logical properties of the relations involved and of the domain of entities to which they are applied to a freely generated set. The effect of such axioms is to shrink the class of semantically distinct relations. Each constraint corresponds to an equivalence relation on the set of freely generated relations. In terms of our generative definitional approach, we can consider these equivalence constraints in the following general cases:

3.2.1. The Case of a Symmetric Primitive Relation

Suppose we start with a symmetric primitive, such as the binary connection relation, **C**, of the RCC theory. We can easily identify the set of Level 1 definable relations of such a relation by considering how symmetry will give rise to logical equivalences among the freely generated Level 1 relations.

As the Table 1 (Column 4) shows for the case of the **C** relation, **0**, **2**, **3** are all equivalent to the empty relation \perp . **1** and **4** are equivalent to the Level 0 relations R and $\neg R$ respectively (**C** and $\neg C$ in the table); so as we stipulated above, they are considered to be Level 0. Furthermore, each of **{5, 9, 13, 17}**, **{6, 10, 14, 18}**, **{7, 11, 15, 19}** and **{8, 12, 16, 20}** is a set of equivalent relations. Thus, symmetry of primitive relation, reduces the 21 relations given above to only 5 distinct Level 1 relations. Since, their negations are also Level 1, the total number of Level 1 relations definable from a symmetric relation is 10.

In constructing Level 2 relations from a symmetric relation, R , we only need consider relations R , **5**, **6** and **8** as constituents, because \perp is not useful and **7** is the converse of **6**.

3.2.2. Reflexivity and Transitivity

Of the most commonly axiomatised properties of relations, symmetry seems to be the one that gives rise to the most significant reduction in the size of the levels in a generative definition survey, especially at the lower levels. Reflexivity and transitivity, do not give rise to any equivalences at levels 0 or 1. However, they do entail subsumptions between relations definable at level 1. Specifically: if R is reflexive then the negated form of **19** is a subset of R (e.g. in the RCC theory we have $P \subseteq C$). If R is transitive then R is a subrelation of R composed with R . These will give rise to significant reductions in the size of levels 2 and above, since they mean that certain intersections will be equivalent to the empty relation.

3.3. Axioms of the Domain as Constraints

In addition to the general properties of the set of primitive relations, we may use a set of axioms to constrain their meaning and the domain of entities to which they apply. This will reduce the set of logically distinct definitions generated.

Allen and Hayes [5] gave an axiomatisation of temporal intervals using the relation meets M as a primitive. For intervals x, y , x meets y , i.e. $M(x, y)$, iff the end of x is simultaneous with the beginning of y . Using these axioms, we have proved that the previous definitions of Allen’s 13 temporal interval relations are equivalent to those that were given in [5]. Moreover, we show in Table 1 (Column 5) that, from the set of freely generated relations at Level 1, the relations **8, 10, 11, 12, 14, 15, 16** and **20** can be inferred from the axioms, being therefore equivalent to the universal relation within this domain. Any instance of relation **1** would be inconsistent with the axioms, so it is equivalent to the empty relation. **2** and **3** are equivalent to the Level 0 relations $M(y, x)$ and $M(x, y)$ respectively. Finally, the pairs **(6, 7)** and **(18, 19)** are equivalent to the negations of **5** and **17** respectively. Therefore, there are only 5 logically distinct new relations and their negations at Level 1. MMi means that either x meets with y or y meets with x .

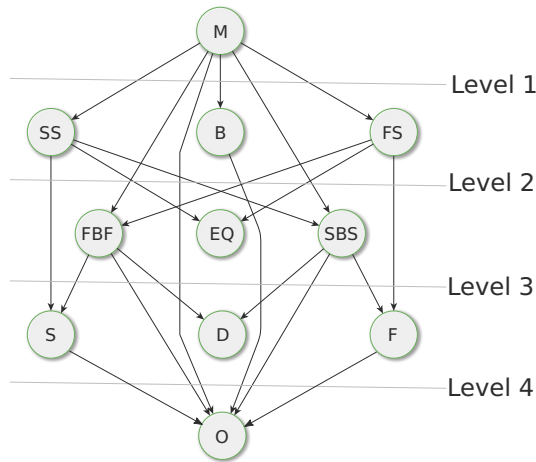


Figure 2. Definition derivation graph for Allen’s relations.

4. Generation of Allen’s 13 Basic Interval Relations

Allen’s Interval Calculus [5] specifies 13 jointly exhaustive and pairwise disjoint (JEPD) relations between temporal intervals: **B** (Before), **M** (Meet), **O** (Overlap), **S** (Start), **D** (During), **F** (Finish); and their inverses: **A** (After), **Mi**, **Oi**, **Si**, **Di**, **Fi**; and **EQ** (Equal). Using M as the primitive relation, Allen’s 13 temporal interval relations can all be defined by 4 or fewer binary relation combination operations. A number of intermediate relations are defined during the constructions: **SS** (Start at Same time), **FS** (Finish at Same time), **SBS**(x, y) (x starts before y starts), **SAS**(x, y) (x starts after y starts), **FBF**(x, y) (x finishes before y finishes), **FAF**(x, y) (x finishes after y finishes) and **BM** (**B** or **M**). Fig. 2 shows the level structure of the following set of definitions:

L0: $M(x, y)$	L3: $S(x, y) \equiv_{def} SS(x, y) \wedge FBF(x, y)$
L0: $Mi(x, y) \equiv_{def} M(y, x)$	L3: $Si(x, y) \equiv_{def} S(y, x)$
L1: $B(x, y) \equiv_{def} \exists z[M(x, z) \wedge M(z, y)]$	L3: $F(x, y) \equiv_{def} SBS(y, x) \wedge FS(x, y)$
L1: $A(x, y) \equiv_{def} B(y, x)$	L3: $Fi(x, y) \equiv_{def} F(y, x)$
L1: $SS(x, y) \equiv_{def} \exists z[M(z, x) \wedge M(z, y)]$	L3: $D(x, y) \equiv_{def} SBS(y, x) \wedge FBF(x, y)$
L1: $FS(x, y) \equiv_{def} \exists z[M(x, z) \wedge M(y, z)]$	L3: $Di(x, y) \equiv_{def} D(y, x)$
L2: $EQ(x, y) \equiv_{def} SS(x, y) \wedge FS(x, y)$	L3: $SBSFBF(x, y) \equiv_{def}$
L2: $SBS(x, y) \equiv_{def} \exists z[SS(x, z) \wedge M(z, y)]$	$SBS(x, y) \wedge FBF(x, y)$
L2: $SAS(x, y) \equiv_{def} SBS(y, x)$	L4: $O(x, y) \equiv_{def} \exists z[S(z, y) \wedge F(z, x)]$
L2: $FBF(x, y) \equiv_{def} \exists z[M(x, z) \wedge FS(z, y)]$	L4: $O(x, y) \equiv_{def}$ (alternatively)
L2: $FAF(x, y) \equiv_{def} FBF(y, x)$	$SBSFBF(x, y) \wedge \neg BM(x, y)$
L2: $BM(x, y) \equiv_{def} B(x, y) \vee M(x, y)$	L4: $Oi(x, y) \equiv_{def} O(y, x)$

5. Region Connection Calculus

The Region Connection Calculus (RCC) [1] is a first-order formalism based on regions and the primitive connection relation, C , axiomatised to be reflexive and symmetric. Two regions x, y are connected ($C(x, y)$ holds), if their closures share a point. Based on this relation, many spatial relations can be defined. Among them, a set of 8 JEPD relations are identified: DC (Disconnected), EC (Externally Connected), PO (Partially Overlap), TPP (Tangential Proper Part), NTPP (Non-Tangential Proper Part), TPPi (Inverse Tangential Proper Part), NTPPi (Inverse Non-Tangential Proper Part) and EQ (Equal). They are referred to as RCC-8, which is well-known in the field of qualitative spatial reasoning. The following 5 JEPD relations are referred to as RCC-5: DR (Discrete), PO, PP (Proper Part), PPI (Inverse Proper Part) and EQ, which is a more granular version of RCC-8, which can be formed by disjoining DC and EC, and each pair of proper part relations. The basic theory can be extended by further definitions to create much more expressive vocabularies of relations (see e.g. [13,19]).

5.1. RCC Level 1 Relations Definable from C

Following the previous discussions and since C is symmetric, Level 1 relations consist of only 5 logically distinct relations and their negations. As shown in Table 1 (Column 4), these are:

L1: $\perp(x, y) \equiv_{def} C(x, y) \wedge \neg C(x, y)$	L1: $JU(x, y) \equiv_{def} \forall z[C(z, x) \vee C(z, y)]$
L1: $P(x, y) \equiv_{def} \forall z[C(z, x) \rightarrow C(z, y)]$	L1: $IndC(x, y) \equiv_{def} \exists z[C(z, x) \wedge C(z, y)]$
L1: $Pi(x, y) \equiv_{def} \forall z[C(z, y) \rightarrow C(z, x)]$	

We may read $JU(x, y)$ as x and y are *jointly universal*. This means that their sum fills the whole space. In other words, everything is connected to x or y (or both). The relation $IndC(x, y)$ is true when x and y are indirectly connected *via* a third region. Within the standard RCC theory $IndC$ is equivalent to the universal relation \top . This is because there is a universal region that is connected to every region. $IndC \equiv \top$ also follows from the fact that, in standard RCC, every pair of regions has a sum, which is connected to both of them. Nevertheless, one can easily imagine theories with a primitive connection relation but a different domain of regions for which $IndC \not\equiv \top$. For instance, the domain could be the nodes of a graph, with $C(x, y)$ meaning that nodes x and y are connected by an edge.

5.2. Level 2 Relations Definable from \mathbf{C}

Assuming that $\text{Ind}\mathbf{C} \equiv \top$, the only relations available for constructing Level 2 relations in RCC are: \mathbf{C} , \mathbf{P} , and \mathbf{JU} . (\mathbf{Pi} is also available, but is a cognate of \mathbf{P} , so need not be considered separately.) The fact that $\mathbf{P} \subseteq \mathbf{C}$ also creates many equivalences. There are still a large number of possibilities. Specifically:

L2: $\mathbf{C}(x, y) \wedge \neg\mathbf{P}(x, y)$	L2: $\mathbf{P}(x, y) \wedge \mathbf{JU}(x, y)$	L2: $\neg\mathbf{P}(y, x) \wedge \mathbf{JU}(x, y)$
L2: $\mathbf{C}(x, y) \wedge \neg\mathbf{P}(y, x)$	L2: $\mathbf{P}(x, y) \wedge \neg\mathbf{JU}(x, y)$	L2: $\neg\mathbf{P}(y, x) \wedge \neg\mathbf{JU}(x, y)$
L2: $\mathbf{C}(x, y) \wedge \mathbf{JU}(x, y)$	L2: $\mathbf{P}(y, x) \wedge \mathbf{JU}(x, y)$	L2: $\mathbf{O}(x, y) \equiv_{\text{def}} \exists z[\mathbf{P}(z, x) \wedge \mathbf{P}(z, y)]$
L2: $\mathbf{C}(x, y) \wedge \neg\mathbf{JU}(x, y)$	L2: $\mathbf{P}(y, x) \wedge \neg\mathbf{JU}(x, y)$	L2: $\text{POCR}(x, y) \equiv_{\text{def}} \exists z[\mathbf{P}(x, z) \wedge \mathbf{P}(y, z)]$
L2: $\neg\mathbf{C}(x, y) \wedge \mathbf{JU}(x, y)$	L2: $\neg\mathbf{P}(x, y) \wedge \mathbf{JU}(x, y)$	
L2: $\neg\mathbf{C}(x, y) \wedge \neg\mathbf{JU}(x, y)$	L2: $\neg\mathbf{P}(x, y) \wedge \neg\mathbf{JU}(x, y)$	

Many of the syntactically possible definitions yield relations equivalent to those at Level 0 or Level 1 and are not shown given here. In the context of the standard interpretation of RCC several of the given relations are not possible, or are equivalent to other relations. For instance, $\neg\mathbf{C}(x, y) \wedge \mathbf{JU}(x, y)$ is not possible unless the universe consists two or more separated components; and POCR ('parts of a common region') would be equivalent to \top as long as every two region has a sum, or there is a universal region, of which every other region is a part.

5.3. Generation of RCC-5 and RCC-8

It turns out that all of the RCC-5 and RCC-8 relations can be generated using simple binary composition and Boolean operators. In fact all RCC-5 relations are generated at level 3 or less and all RCC-8 at level 4 or less:

L0: $\mathbf{C}(x, y)$	(Primitive)
L0: $\text{DC} \equiv_{\text{def}} \neg\mathbf{C}(x, y)$	(RCC-8)
L1: $\mathbf{P}(x, y) \equiv_{\text{def}} \forall z[\mathbf{C}(z, x) \rightarrow \mathbf{C}(z, y)]$	
L1: $\mathbf{Pi}(x, y) \equiv_{\text{def}} \mathbf{P}(y, x)$	
L2: $\text{EQ}(x, y) \equiv_{\text{def}} \mathbf{P}(x, y) \wedge \mathbf{P}(y, x)$	(RCC-5, RCC-8)
L2: $\text{PP}(x, y) \equiv_{\text{def}} \mathbf{P}(x, y) \wedge \neg\mathbf{P}(y, x)$	(RCC-5)
L2: $\text{PPi}(x, y) \equiv_{\text{def}} \text{PP}(y, x)$	(RCC-5)
L2: $\text{P-SC} \equiv_{\text{def}} \mathbf{P}(x, y) \vee \mathbf{P}(y, x)$	(the symmetric closure of \mathbf{P})
L2: $\mathbf{O}(x, y) \equiv_{\text{def}} \exists z[\mathbf{P}(z, x) \wedge \mathbf{P}(z, y)]$	
L2: $\text{DR}(x, y) \equiv_{\text{def}} \neg\mathbf{O}(x, y)$	(RCC-5)
L3: $\text{EC}(x, y) \equiv_{\text{def}} \mathbf{C}(x, y) \wedge \neg\mathbf{O}(x, y)$	(RCC-8)
L3: $\text{PO}(x, y) \equiv_{\text{def}} \mathbf{O}(x, y) \wedge \neg\text{P-SC}(x, y)$	(RCC-5, RCC-8)
L3: $\text{NTPP}(x, y) \equiv_{\text{def}} \neg\exists z[\mathbf{C}(x, z) \wedge \text{DR}(z, y)]$	(RCC-8)
L3: $\text{NTPPi}(x, y) \equiv_{\text{def}} \text{NTPP}(y, x)$	(RCC-8)
L4: $\text{TPP}(x, y) \equiv_{\text{def}} \text{PP}(x, y) \wedge \neg\text{NTPP}(x, y)$	(RCC-8)
L4: $\text{TPPi}(x, y) \equiv_{\text{def}} \text{TPP}(y, x)$	(RCC-8)

Our **L3** definition of NTPP is considerably simpler than the previously published **L5** definition, $\text{NTPP}(x, y) \equiv_{\text{def}} \text{PP}(x, y) \wedge \neg\exists z[\text{EC}(z, x) \wedge \text{EC}(z, y)]$ [1].

5.4. Defining Complement

This systematic strategy enables the selection of a definition at its lowest possible level, given that, through the generation process, a definition is only added to a level if there are

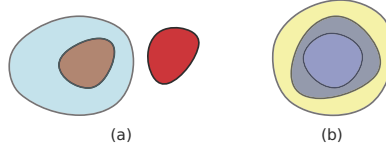


Figure 3. Cases of *non-tangential partial overlap* (NTPO). In (a) a red region is in two disjoint parts, one is part of and the other is disjoint from the cyan region. In (b) the yellow region has a hole, which is within the area covered by the blue region.

no prior equivalences. As an example, consider the definition of $\text{Compl}(x, y)$, meaning that x is the complement of y (i.e. the largest region that does not overlap y). The early RCC paper [1] contains a definition of a complement function, which implicitly defines a complement relation by:⁴

$$\mathbf{L5: Compl}(x, y) \equiv_{def} \forall z [\mathbf{C}(z, x) \leftrightarrow \neg \mathbf{NTPP}(z, y)] \wedge \mathbf{O}(z, x) \leftrightarrow \neg \mathbf{P}(x, y)]$$

However, while searching the space of possible definitions, we found the following much simpler definition at level 3:

$$\mathbf{L3: Compl}(x, y) \equiv_{def} \mathbf{JU}(x, y) \wedge \neg \mathbf{O}(x, y)$$

5.5. Case Study of the NTPO Relation

The binary relation combining rules provide a systematic way of generating more complex relations from existing sets. The following sequence of definitions leads up to the definition of *non-tangential partial overlap*, NTPO, which occurs when the interior and exterior of one region both have intersections with the interior and exterior of another region, but boundaries of these two regions do not intersect. In other words, each region has a part within and a part outside the other, but their boundaries do not intersect. Two examples are illustrated in Figure 3.

$$\mathbf{L4: NTPP-SC}(x, y) \equiv_{def} \mathbf{NTPP}(x, y) \vee \mathbf{NTPP}(y, x)$$

$$\mathbf{L5: Compo+}(x, y) \equiv_{def} \forall z [\mathbf{TPP}(z, x) \rightarrow \mathbf{TPP}(z, y)]$$

$$\mathbf{L6: R1}(x, y) \equiv_{def} \exists z [\mathbf{PP}(z, x) \wedge \mathbf{Compo+}(z, y)]$$

$$\mathbf{L7: Compo}(x, y) \equiv_{def} \mathbf{Compo+}(x, y) \wedge \neg \mathbf{R1}(x, y)$$

$$\mathbf{L5: SDCB}(x, y) \equiv_{def} \mathbf{DC}(x, y) \vee \mathbf{NTPP-SC}(x, y)$$

$$\mathbf{L8: R2}(x, y) \equiv_{def} \forall z [\mathbf{Compo}(z, x) \rightarrow \mathbf{SDCB}(z, y)]$$

$$\mathbf{L9: R3}(x, y) \equiv_{def} \forall z [\mathbf{Compo}(z, x) \rightarrow \mathbf{R2}(y, z)]$$

$$\mathbf{L10: R4}(x, y) \equiv_{def} \exists z [\mathbf{Compl}(x, z) \wedge \mathbf{R3}(z, y)]$$

$$\mathbf{L11: R5}(x, y) \equiv_{def} \exists z [\mathbf{Compl}(x, z) \wedge \mathbf{R4}(y, z)]$$

$$\mathbf{L13: R6}(x, y) \equiv_{def} \mathbf{R3}(x, y) \vee \mathbf{R4}(x, y) \vee \mathbf{R4}(y, x) \vee \mathbf{R5}(x, y)$$

$$\mathbf{L14: NTPO4}(x, y) \equiv_{def} \mathbf{PO}(x, y) \wedge \mathbf{R6}(x, y)$$

⁴It is not entirely clear why this definition has both of these conjuncts. It seems that the definition also has axiomatic import in that it constrains the relationship between \mathbf{C} and \mathbf{O} .



Figure 4. (a) illustrates different types of component of a region. (b) Illustrates the ‘straddling’ relation. In this case the red region straddles (the boundary of) the light blue region.

$\text{Compo}(x, y)$ means ‘ x is a component of y ’ — i.e. a maximal self connected part (or whole piece). A multi-piece region is equal to the sum of its components. Note that $\text{Compo}(x, x)$ holds just in case x is self connected (i.e. in a single piece). The definition of $\text{Compo}(x, y)$ also allows the case of a component that is connected by a one or more single points (in 2D) or line-like contacts (in 3D), as shown in Fig. 4(a). However, these different types of component can be distinguished by further definitions. $\text{Compo}^+(x, y)$ is read ‘ x is a componential part of y ’ and means that x is a part of y such that x is the sum of one or more (possibly all) components of y . This means that x is the sum of any number of self connected parts of y . $\text{R1}(x, y)$ holds if x has a *proper* part that is a component of y . It is used in combination with Compo^+ to define Compo , since a single component cannot have a sum of components as a proper part.

NTPP-SC is the *symmetric closure* of NTPP . $\text{SDCB}(x, y)$ may be read as ‘ x and y are in a *Simple DisConnected Boundary relationship*’. This holds when either x and y are disconnected or one is a non-tangential proper part of the other. These are ‘simple’ cases where the boundaries of x and y do not intersect. However, this relation does not cover all cases where boundaries do not intersect, because SDCB requires that all components of region x stand in the same topological relation to region y (and *vice versa*). In the case where both regions are single component regions with no holes, they can have disconnected boundaries only if the relation SDCB holds. But if either region is multi-piece or has holes then there are more cases that SDCB does not cover, and will be explained below.

$\text{R2}(x, y)$ holds when all components of x stand in relation SDCB to y . This is a generalisation of $\text{SDCB}(x, y)$, since it allows different components of x to stand in different topological relations to y . Expanding R2 in the definition of R3 , we have: $\forall z_1[\text{Compo}(z_1, x) \rightarrow [\forall z_2[\text{Compo}(z_2, y) \rightarrow \text{SDCB}(z_2, z_1)]]]$, which is equivalent to $\forall z_1 \forall z_2[\text{Compo}(z_1, x) \wedge \text{Compo}(z_2, y) \rightarrow \text{SDCB}(z_1, z_2)]$. Hence, $\text{R3}(x, y)$ holds if for every component z_1 of x and for every component z_2 of y , $\text{SDCB}(z_1, z_2)$ holds. $\text{R3}(x, y)$ is a symmetric relation.

R4 and R5 are variants of R3 where we consider the relationship involving components of the regions’ complements. So $\text{R4}(x, y)$ holds if R3 holds between the complements of x and y . The definition of R5 can be rewritten as $\exists z_1[\text{Compl}(x, z_1) \wedge [\exists z_2[\text{Compl}(y, z_2) \wedge \text{R3}(z_2, z_1)]]]$ This is equivalent to: $\exists z_1 \exists z_2[\text{Compl}(x, z_1) \wedge \text{Compl}(y, z_2) \wedge \text{R3}(z_2, z_1)]$. Hence, $\text{R5}(x, y)$ holds if R3 holds between the complement of x and the complement of y . R5 is a symmetric relation.

$\text{R6}(x, y)$ specifies 4 different cases where the boundaries of x and y do not intersect. $\text{NTPO4}(x, y)$ is a specialisation of $\text{PO}(x, y)$ such that $\text{R6}(x, y)$ holds additionally. It is easy to see that $\text{R6}(x, y)$ is a symmetric relation, thus $\text{NTPO4}(x, y)$ is also symmetric. However, $\text{NTPO4}(x, y)$ still does not cover all the cases where x and y partially overlaps but do not have connected boundaries. $\text{NTPO4}(x, y)$ can be extended further as follows. For every component x_1 of x , we take the complement of it as x_2 , then every component x_3 of x_2 must have a connected boundary. Let us call x_3 a CCComponent of

x . $\text{NTPO}(x, y)$ is defined regarding their CCComponent s. Let $\text{R7}(x, y)$ be a relation which holds iff for every CCComponent x_3 of x , for every CCComponent y_3 of y , such that $\text{SDCB}(x_3, y_3)$ holds. Then $\text{NTPO}(x, y) \equiv_{\text{def}} \text{PO}(x, y) \wedge \text{R7}(x, y)$. By definition, $\text{NTPO}(x, y)$ is symmetric. Below, we show that R7 is definable in a similar manner to R3 and R5 .

- L9:** $\text{R2c}(x, y) \equiv_{\text{def}} \exists z[\text{Compl}(x, z) \wedge \text{R2}(z, y)]$
L10: $\text{R3c}(x, y) \equiv_{\text{def}} \forall z[\text{Compo}(z, x) \rightarrow \text{R2c}(z, y)]$
L11: $\text{R4c}(x, y) \equiv_{\text{def}} \forall z[\text{Compo}(z, x) \rightarrow \text{R3c}(y, z)]$
L12: $\text{R5c}(x, y) \equiv_{\text{def}} \exists z[\text{Compl}(x, z) \wedge \text{R4c}(z, y)]$
L13: $\text{R7}(x, y) \equiv_{\text{def}} \forall z[\text{Compo}(z, x) \rightarrow \text{R5c}(z, y)]$
L14: $\text{NTPO}(x, y) \equiv_{\text{def}} \text{PO}(x, y) \wedge \text{R7}(x, y)$

NTPO seems to be a natural specialisation of PO ; however, a long sequence of binary definitions is required to specify it. The complexity of the definition seems to arise because of the many topologically distinct ways in which NTPO can occur. Our investigation of the construction of the NTPO definition according to our binary combination rules has revealed simpler definitions of the component relationships than had been previously given. However, due to the huge proliferation of possible relation definitions from C at higher levels, we have not yet been able to exhaustively search for the shortest possible definition of NTPO . There may well be a much more straightforward definition (perhaps *via* the ‘straddling’ relation shown in Fig. 4).

Indeed, if we allow the *conjoined composition* operation that was mentioned in Section 2.2, we can generate NTPO much more directly. First we define a boundary connectedness relation $\text{BC}(x, y)$ by the following disjunction of conjoined compositions (arrows denote the directions of non-symmetric relations, e.g. TP):

$$\text{BC}(x, y) \equiv_{\text{def}} \begin{array}{c} \begin{array}{ccc} & Z_1 & \\ \text{TPPi} \nearrow & \text{C} & \searrow \text{TPP} \\ X & & Y \\ \text{EC} \searrow & & \nearrow \text{EC} \\ & Z_2 & \end{array} \vee \begin{array}{ccc} & Z_1 & \\ \text{TPPi} \nearrow & \text{C} & \searrow \text{EC} \\ X & & Y \\ \text{EC} \searrow & & \nearrow \text{TPP} \\ & Z_2 & \end{array} \vee \text{EQ}(x, y) \end{array}$$

The two disjuncts correspond to the case where x and y share a common part of their boundary either from the same or opposite side of their boundaries. We then define $\text{NTPO}(x, y) \equiv_{\text{def}} \text{PO}(x, y) \wedge \neg \text{BC}(x, y)$.

6. Level-by-Level Analysis of Allen, RCC and Other Relation Sets

The work done within the previous sections enables an analysis and comparison of both relation sets. Although the number of logically distinct relations on the first level is similar, the two relation sets behave differently in the generations of further levels. While Allen’s distinct relations are limited to the 13 defined relations and their disjunctions [5] [20], RCC definitions can increase in complexity rapidly. This is due to the fact that the five axioms of the Allen’s theory restrict heavily the number of possible relations. In practice, they define a domain of time intervals which are convex and linear on time. On the other hand, RCC ’s domain is composed by convex and concave regions within multiple dimensions [19].

With automated generation of the Allen relations, the process of filtering semantically distinct relations from the freely generated set is hard because we need to compute equivalences with respect to rather complex axioms. However, this process will end at level 6, providing a finite set of relations. By contrast, the generation of RCC levels is initially much easier because the RCC axioms of symmetry and reflexivity enforce strong constraints by means of simple axioms, but the number of distinct relations grows increasingly through generations.

In future development of this kind of definitional analysis, we suggest the notion of a *generative definitional survey*. This would be an exhaustive listing (relative to some precise specification of definitional level) of all relations definable up to a given definitional level, from a given primitive relation (or set of relations) in the presence of certain axiomatically specified constraints. The axiomatic constraints act as a filter, which reduces the number of semantically distinct relations at each level. A generative definitional survey may be regarded as a representation of compiled semantic knowledge. As such, it is similar to the *composition tables* that are often used to facilitate reasoning with relational calculi. In fact, because composition is one of our definition construction operators, the composition table for any set of relations may be directly obtained from a sufficiently deep generative definitional survey that includes those relations.

We have some initial results regarding definitional surveys up to Level 2, using standard composition and the binary combination operators. Our approach is to first generate the freely definable relations of a level using Prolog, then identify equivalences using a combination of Mace4 and Prover9 [21]. Although, no longer state of the art, these are certainly sufficient to derive the first two levels arising with a number of commonly applied constraints. We initially attempt to find a model using Mace to check if a counter-example to the equivalence can be found; if not, we try to prove equivalence using Prover9. This method seems to be very effective. Our initial results are as follows:

Level	free	sym	sym+ref	Allen
0	4	2	2	4
1	42	10	10	10
2	5326	242	216	

7. Summary and Further Work

In this paper we have considered the problem of generating relational calculi from a set of primitive relations, of which RCC and Interval Algebra are two well known examples. Whereas these calculi originally employed definitions constructed in an *ad hoc* way, we have shown that such definitions be generated in a more systematic fashion. It is interesting to compare this approach against the n -intersection framework promulgated by Egenhofer and his colleagues [22], which starts off with a fixed set of possible relations (2^9 in the 9-intersection case) and then reduces this to a much smaller set by imposing various conditions on the regions. Here we start off with a very small number of relations (the primitives) and progressively discover more and more new relations, potentially without bound.

There are a variety of directions in which this work could be taken further. Firstly, other relational calculi could be investigated, for example, the INDU calculus which extends the Interval Algebra [23]. Secondly, the approach could be applied to other

definitional frameworks, for example relational algebra (in which RCC has also been defined [18]), or to Description Logic [24].

References

- [1] D. A. Randell, Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. In *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, pages 165–176, 1992.
- [2] B. Bennett, A. Isli, and A. G. Cohn. When does a composition table provide a complete and tractable proof procedure for a relational constraint language? In *Proceedings of the IJCAI-97 workshop on spatial and temporal reasoning*, Nagoya, Japan, 1997.
- [3] J. Renz and B. Nebel. On the complexity of qualitative spatial reasoning: a maximal tractable fragment of the Region Connection Calculus. In *Proceedings of IJCAI-97*, 1997.
- [4] E. Davis, N. Gotts, and A. G. Cohn. Constraint networks of topological relations and convexity. *Constraints*, 4(3):241–280, 1999.
- [5] J. F. Allen and P. J. Hayes. A common sense theory of time. In *Proceedings 9th IJCAI*, pages 528–531, Los Angeles, USA, 1985.
- [6] S. Borgo, N. Guarino, and C. Masolo. A pointless theory of space based on strong congruence and connection. In *Principles of Knowledge Representation and Reasoning, Proceedings of KR-96*, 1996.
- [7] B. Bennett. A categorical axiomatisation of region-based geometry. *Fundamenta Informaticae*, 46(1–2):145–158, 2001.
- [8] T. Laguna. Point, line and surface as sets of solids. *Journal of Philosophy*, 19:449–61, 1922.
- [9] A. Tarski. Foundations of the geometry of solids. In *Logic, Semantics, Metamathematics*, chapter 2. Oxford Clarendon Press, 1956. trans. J.H. Woodger.
- [10] L. Henkin, P. Suppes, and A. Tarski. *The Axiomatic Method — with special reference to geometry and physics*. North-Holland, Amsterdam, 1959.
- [11] R. M. Robinson. Binary relations as primitive notions in elementary geometry. In *The Axiomatic Method (with special reference to geometry and physics)*, pages 68–85, 1959.
- [12] A. Tarski. Some methodological investigations on the definability of concepts. In *Logic, Semantics, Metamathematics*. Oxford Clarendon Press, 1956. trans. J.H. Woodger.
- [13] A. G. Cohn, B. Bennett, J. Gooday, and N. Gotts. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1:275–316, 1997.
- [14] M. Egenhofer. Reasoning about binary topological relations. In *Proceedings of the Second Symposium on Large Spatial Databases, SSD'91 (Zurich)*, pages 143–160, 1991.
- [15] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6:73–89, 1941.
- [16] A. Tarski and S. Givant. *A Formalization of Set Theory without Variables*. Number 41 in Colloquium Publications. American Mathematical Society, Rhode Island, 1987.
- [17] P. Ladkin and R. Maddux. On binary constraint problems. *Journal of the ACM*, 41(3):435–469, 1994.
- [18] I. Düntsch and E. Orłowska. A proof system for contact relation algebras. *Journal of Philosophical Logic*, 29:241–262, 2000.
- [19] N. M. Gotts. How far can we ‘C’? defining a ‘doughnut’ using connection alone. In *Principles of KR: Proceedings of the 4th International Conference (KR94)*. Morgan Kaufmann, 1994.
- [20] P. Ladkin and R. Maddux. *Representation and reasoning with convex time intervals*. 1988.
- [21] Prover9 and Mace4. <https://www.cs.unm.edu/~mccune/mace4/>. Accessed 1st May 2016.
- [22] M. Egenhofer and R. Franzosa. Point-set topological spatial relations. *International Journal of Geographic Information Systems*, 5(2):161–174, 1991.
- [23] A. Pujari, G. Kumari, and A. Sattar. INDU: an interval and duration network. In *Advanced Topics in Artificial Intelligence*, pages 291–303. Springer, 1999.
- [24] D. Calvanese, M. Lenzerini, and D. Nardi. Description logics for conceptual data modelling. In *Logics for Databases and Information Systems*, chapter 1. 1998.