

Computing Human-Like Qualitative Topological Relations via Visual Routines

Andrew Lovett

Kenneth Forbus

Qualitative Reasoning Group, Northwestern University
2133 Sheridan Rd, Evanston, IL 60208 USA
Andrew-Lovett@northwestern.edu
Forbus@northwestern.edu

Abstract

A core problem in spatial reasoning is finding an appropriate set of relationships to compute. This paper proposes that humans represent topological relationships between 2D regions using three basic, qualitative relations: **contains**, **intersects**, and **overlaps-with**. We show how these relations can be computed from sketched inputs using a model of mid-level perception. Results from a pilot experiment indicate that these three relationships suffice to explain people’s judgments on four English spatial terms (“intersects”, “overlaps”, “connects to”, and “contains”), although a combination of the three is generally required for each term.

Introduction

A major problem in building systems that reason about space is determining the correct set of spatial relations to represent. In the QR community, the Region Connection Calculus (RCC8) (Cohn 1996; Cohn et al. 1997; see Figure 1) is a prominent and effective way of representing topological relations between two-dimensional shapes. RCC8 includes 8 qualitative terms which exhaustively describe the set of possible topological relations between two shapes. RCC8 relations have been used in a number of applications, from qualitative spatial simulation (Randell et al. 1992) to sketch understanding (Forbus et al. 2008).

While representational schemes like RCC8 are useful for building formal AI reasoning systems, it is not clear how closely they align with human spatial representations. Reasoning systems which use human-like representations are better equipped for both interacting with humans in cooperative endeavors and modeling human thought processes in cognitive modeling studies. However, there have been few attempts by AI researchers to look at how humans compute and represent topological relations.

In one notable exception from Geographic Information Systems, Xu and Mark (1997) conducted a study in which they showed participants scenes containing pairs of linear objects (such as roads and rivers). Participants were instructed to indicate how well various predicates

described the scenes (predicates included “X crosses Y,” “X connects with Y,” “X merges with Y,” etc). By studying their results, the authors were able to get a better idea of the various factors that determined which predicate people might use in describing a geographic scene.

While the Xu and Mark results are helpful, we believe there is a more general question of what are the topological primitives computed and represented by humans when they examine a visual scene. By *primitives*, we mean a small set of relations from which all (or at least most) other topological relations can be computed. These primitives should meet the following requirements:

- 1) They should be easily computable by humans using low- or mid-level visual operations.
- 2) They should not be tied to any particular domain, such as geography.
- 3) While the individual primitives may not correspond to topological terms in the English language, such as “contains” or “intersects with,” it should be possible to explain how humans can use the primitives together to compute and assess those terms.

In this paper, we propose that people use three topological primitives for representing two-dimensional visual scenes: **contains**, **intersects**, and **overlaps-with**. We show how these primitives can be computed using *visual routines* (Ullman 1984), a general approach to modeling

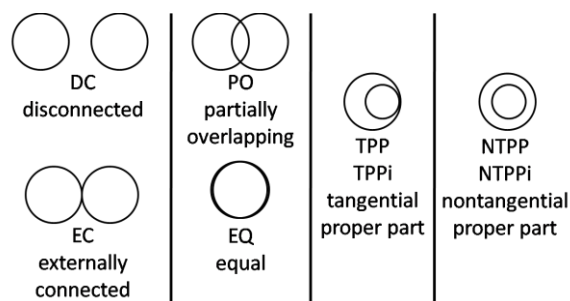


Figure 1. The Region Connection Calculus (RCC8) relations for describing topology. The TPP and NTPP relations each have inverse relations, TPPI and NTPPI.

Operation Type	Operations
Covert Attention	Curve Tracing, Scanning, Region Coloring
Working with Elements or Objects	Attribute Access, Activation, Inhibition/Excitation, Deletion
Working with Objects	Object Creation, Binding to Elements
Maintenance	Marking Locations

Table 1. Elementary operations in VRS.

mid-level visual processing. We then evaluate the primitives by examining how well they explain human assessments of four topological terms from English: “connects,” “intersects,” “overlaps,” and “contains.” Note that while the primitives and the English terms look quite similar, our results show that there is by no means a one-to-one mapping between primitives and English terms.

We begin by presenting Visual Routines for Sketching (VRS), an implementation of Ullman’s visual routines proposal which we are developing. We then summarize the psychological literature on topological relations and show how it motivates the use of our three topological primitives. Then, we describe the visual routines written in VRS to compute our three primitives. After this, we present the results from a preliminary psychological study conducted to evaluate our primitives. We conclude by discussing related and future work.

Visual Routines for Sketching

Ullman (1984) proposed that people have access to a set of *elementary operations*, operations we can run over our visual working memory to extract information. This finite set of operations can be combined in different ways to create a near-infinite set of *visual routines* for computing different spatial features and relations.

We are developing Visual Routines for Sketching (VRS), a computer implementation of visual routines, as a platform for experimenting with computational models of perception. It provides a set of low-level elementary operations, supported by the psychophysics and cognitive psychology literature. Using these operations, researchers can construct visual routines based on their theories for how a particular spatial feature is computed. These routines can be run and evaluated on two-dimensional sketches or line drawings in CogSketch¹ (Forbus et al. 2008), an open-domain sketch understanding system.

CogSketch users can create sketches either by drawing with a pen or by importing shapes built in PowerPoint. VRS works directly with the *ink* of the sketch, the lines representing the edges of each object. Thus, it avoids edge detection issues.

VRS’s current vocabulary of operations is given in Table 1. As we describe each of the levels of representation in the system, we refer to operations listed in this table.

Basic Representation

Ullman (1984) suggested that the human perceptual system uses a bottom-up, parallel approach to build an initial *basic representation* of the visual world. VRS computes a basic representation via two steps: First, the ink is projected onto a retinotopic map, a simplification of V1 in the primary visual cortex which represents the orientation of any edges at each location in the image. This produces a set of *edge activations* at various locations.

Second, edge activations are grouped together to form *contours*. This step is based on the contour integration literature (e.g., Yen and Finkel 1998; Li 1998), which suggests that there is a parallel process in which people group edges together based on the Gestalt grouping principles of good continuation and closedness. To these principles we add the hard constraint of uniform connectedness (Palmer and Rock, 1994). That is, edge activations will only be grouped together in a contour if they are the same color and they lie directly adjacent to each other in the visual representation. In the future, we plan to relax the connectedness constraint partially to allow the system fill in gaps between parts of a line (e.g., Saund, 2003).

Incremental Representation

Ullman proposed that there is a set of elementary operations that can be applied serially to the basic representation. By combining these operations into visual routines, an individual can both gather information and update the representation, thus producing an *incremental representation*. In VRS there are three key elementary operations, inspired by Ullman’s proposal, which gather data and add *visual elements* to the incremental representation:

1) *Curve Tracing* traces along consecutive edge activations. It produces a *curve*, a new grouping of activations which may lie along one or multiple contours.

2) *Scanning* begins at one location and moves forward in a fixed direction. It produces a straight curve representing the line scanned over.

3) *Region Coloring* fills in the area between curves and contours, creating a new *region*.

All three operations take optional arguments that allow them to be constrained in several ways, e.g., curve tracing along a region, region coloring along a curve, or scanning between two points. The operations can be used to gather information, such as detecting what other elements lie along a curve or within a region.

¹ Available for download at:
http://silccenter.org/projects/cogsketch_index.html

The visual elements in the incremental representation can be queried via the *Attribute Access* operation to access data such as the size of an element, the center of an element, the curvedness of a curve, or the orientation of a straight curve. Elements can also be *Inhibited*, causing them to be ignored by future operations.

Objects

The *Object Creation* operation sets up *object files* (Kahneman et al., 1992). Object files serve as a bridge between the visual representation and higher-level, conceptual representations. Each object file contains indices (Pylyshyn, 2001), which point to the curves and regions that make up the object in the incremental representation. Because objects can share regions or curves in the incremental representation (as when two shapes overlap), it is possible for multiple object files to point down to the same visual elements in the incremental representation. However, these elements can only point up to one object file at a time. To ensure that a particular object file's visual elements are pointing up to it, a routine must *Activate* that object.

Universal Routine

Different visual routines may be relevant to studying different images. However, there needs to be some type of routine to run on the basic representation and gain enough information to determine what follow-up routine to use. Thus, Ullman suggested that there might be a *universal routine* which is applied by default to visual stimuli. The following is a universal routine written using the elementary operations described above. This routine identifies the objects in a visual scene.

Universal Routine: Finding objects in the visual scene
1) Region Coloring: Color the ground, locate any contours in it.
2) Curve Tracing: Trace each contour to determine whether it is a closed shape. Produces a curve.
3) Object Creation: Make an object file for each curve.
4) Region Coloring: If an object is a closed shape, color the area inside it to identify its interior. Produces one or more regions, which will be bound to the object. May also locate new contours located within the object.
5) Recursion: For any new contours located, repeat steps 2-5.

Current State of VRS

At present, VRS contains the elementary operations listed in Table 1. However, we are still in the process of determining the full set of operations and the ways they can interact. Eventually, we hope to develop a simple coding language which will allow other researchers to build their own visual routines by combining elementary operations in novel ways.

Psychological Motivation

Much of the psychological work on topological relations has been related to linguistic terms and how they vary across languages and cultures. Landau and Jackendoff (1993) analyzed the full set of spatial prepositions in the English language—several of which describe topological relations—and determined the various factors that determined which preposition is used to describe a scenario. One important factor was distance. Different distances resulted in the use of different prepositions for describing the relative positions of two objects:

Inside: “in,” “inside,” “throughout”

Contact: “on,” “all over”

Proximal: “near,” “all around”

Distal: “far”

Here, both **inside** and **contact** could be seen as topological primitives that determine which preposition should be used. Landau and Jackendoff further found that the preposition used was only rarely affected by the form of the objects being related to each other. They suggested that our mental representations of relative location are separate from our representations of shape and identity, and they predicted that other languages would similarly use spatial prepositions that were not related to the objects' forms.

A number of studies have found fault with this prediction (see Kemmerer 2006 for a review). There are languages that base the preposition used on the form of the objects being related (e.g., relative tightness of an object in a container for Korean: Hespos and Spelke 2004).

However, there may still be some set of domain-independent topological primitives that are universally computed. These primitives might be combined with object shape and object identity in determining which spatial preposition should be used, with the appropriate combinations varying across languages. Levinson and Meira (2003) conducted a survey of nine highly different languages in which speakers of each language were shown the same set of pictures depicting topological relations and asked to describe those pictures. While there were major differences in how each language grouped the pictures, there appeared to be correlations across languages. Multidimensional scaling revealed that many languages group together pictures relating to *in* (e.g., an animal in a cage), *attachment* (clothes on a clothesline), *on/over* (an object on a table), *on-top* (a tablecloth covering a table), and *near/under*. These groups align with the commonly discussed topological concepts of *containment* and *attachment*, and the physical concept of *support*.

While the distinction between these concepts clearly depends upon the forms of the objects being related, and the distinctions tend to be even more fine grained in many languages, it seems reasonable to propose that Landau and Jackendoff's primitives, **inside** and **contact**, likely aid in distinguishing between *containment*, when one object is

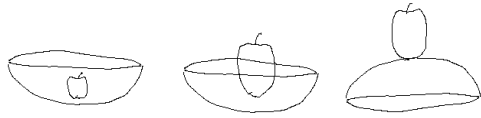


Figure 2. From left to right, the apple is **inside** the bowl, the apple **overlaps** with the bowl, and the apple **intersects** the bowl.

located entirely within another object, and *attachment* or *support*, when the objects are merely touching.

However, we believe that these two primitives are not sufficiently detailed. There are multiple possible forms of contact between two objects in a visual scene. In the simplest form, **intersection**, the edges of the objects simply touch each other in some way. In another form, **overlap**, there is space in the visual scene which is occupied by both the objects. For example, in Figure 2 both the apple inside the bowl and the apple that overlaps with the bowl would be labeled as “in the bowl,” whereas the apple that merely intersects the bowl would be labeled as “on the bowl.” In this paper, we will be testing the hypothesis that people use both the **intersection** and **overlap** primitives, along with **containment**, to compute and assess topological relations.

The Primitives

We have chosen to use three topological primitives: **contains**, **intersects**, and **overlaps-with**. Each of these primitives describes the location of one object, the *target*, relative to another target, the *ground*. In this section, we describe what these primitives mean and give the visual routines for computing them. All visual routines are computed over objects which can be identified in the visual scene using the universal routine described above.

Intersects

This relationship holds whenever some part of one shape’s edge intersects some part of the other shape’s edge. The visual routine for computing this is given below.

Intersects (Target, Referent)

- 1) **Activation**: Activate the Referent object, causing all its associated edges to point up to it.
- 2) **Curve Tracing**: Trace along the Target object’s curve, checking whether any of the Referent object’s edges are encountered.

Overlaps-with

This relation is defined only for pairs of closed shapes (although variations might apply to other shape types). Two shapes are overlapping if their interiors share some region. That is, there is some area that lies within both closed shapes. However, the shapes must also both have regions that are not shared: one shape cannot lie entirely inside the other shape. Note that if one shape **overlaps-**

with another shape, it necessarily also **intersects** the other shape. The visual routine is given below.

Overlaps-with (Target, Referent)

- 1) **Attribute Access**: Check whether the Referent and Target objects share any regions.
- 2) **Attribute Access**: Check whether the Referent has regions not shared by the Target.
- 3) **Attribute Access**: Check whether the Target has regions not shared by the Referent.
- 4) **Combine Data**: If the objects share regions but they both have regions not shared with the other, then they overlap.

Contains

This relation is defined only when one object, the referent, is a closed shape. **Contains** holds when the other object, the target, lies entirely within the referent. The visual routine is given below. Note that this routine actually calls the **overlaps** routine.

Contains (Referent, Target)

- 1) **Activation**: Activate the Target object, causing all its associated edges to point up to it.
- 2) **Region Coloring**: Color in the Referent’s regions, checking to see whether any of the Target’s edges lie within the Referent.
- 3) **Visual Routine Call**: Check whether **Overlaps-with**(Target, Referent) is false.
- 4) **Combine Data**: If part of the Target lies within the Referent, and the Target and Referent do not overlap, then the Referent contains the Target.

Relation to RCC8

Recall that RCC8 (Figure 1) consists of six topological relations, plus two inverse relations, whereas our approach uses only three relations. Nonetheless, all of the RCC8 relations except EQ^2 (equal) can be easily computed from our three relations (see Table 2). We believe this supports our argument that our relations are more basic, or more fundamental. In particular, RCC8 distinguishes between “Tangential Proper Part” (TPP) and “Non-Tangential Proper Part” (NTPP). It seems unlikely that humans make this distinction, at least in their initial representations. The more primitive **contains** relationship captures the important commonalities across TPP and NTPP.

Experiment

We conducted a pilot psychological study to evaluate our primitives. In this study, participants saw basic visual

² It would be relatively straightforward to write a visual routine to compute the **equal** relationship. However, we think it unlikely that humans encode such a relationship, since two objects whose regions and edges are identical will be indistinguishable from each other.

RCC8 Relation	Primitives to Compute it
DC	!intersects ^ !contains
EC	intersects ^ !contains ^ !overlaps
PO	overlaps
TPP/TPPi	contains ^ intersects
NTPP/NTPPi	contains ^ !intersects

Table 2. The topological primitives which can be used to compute seven of the eight RCC8 relations.

scenes consisting of a large red circle and a small green circle (see Figure 3). These scenes were accompanied by a statement such as “Red intersects with green.” Participants were instructed to rate the appropriateness of the statement as a description of the scene, using a scale from 0 to 10.

We evaluated our model by examining whether the topological primitives **contains**, **intersects**, and **overlaps-with** could explain individuals’ ratings for English terms. We assume that an individual might assess a statement such as “Red overlaps with green” by computing some linear combination of the three primitives.

We hypothesized that, if our model was accurate, it should explain both average and individual performance. That is, (1) For each English term, there should be a set of weights for the primitives that correlates highly with average human ratings for that term. This means that the weights are expressing the degree to which individuals consider each of the primitives *on average* in assessing that term. (2) For each English term and each participant, there should be a set of weights for the primitives that correlates highly with that individual’s ratings. This set of weights describes what that particular person considers when assessing the English term. Note that there might be high inter-individual differences in the weights. However, if all individuals are basing their assessments on the primitives, then there should be some appropriate set of weights for all individuals.

Methods

Stimuli consisted of a red circle with radius .5 inches and a green circle with radius .2 inches. There were nine possible distances between the green circle and the red circle, which varied from the two circles being entirely disconnected to the circles overlapping to the green circle being located entirely within the red circle (see Figure 3). There were also four possible directions between the red circle’s center and the green circle’s center (up, down, left, and right). Thus, there were 36 total images.

Each image was accompanied by one of the following sentences:

“Red intersects green.”

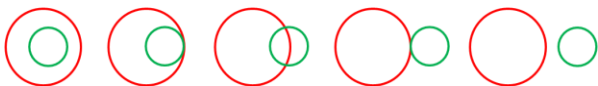


Figure 3. Five of the nine total distances between the large red and small green circles.

“Red overlaps with green.”

“Red connects to green.”

“Red contains green.”

A given participant saw each sentence paired with each image, for a total of $36 \times 4 = 144$ trials. The trials were presented in a random order for each participant.

Participants chose a rating from 0 to 10 for each image/sentence pair by selecting a value from a pop-out menu. Participants were given as much time as they desired to choose the ratings. However, participants chose ratings relatively fast, typically going through the 144 trials in about ten minutes.

The study was run using 10 participants, five male and five female. Nine spoke English as a first language, while the other had learned English at an early age.

Analysis

Our primary question in analyzing the results was whether participants’ ratings could be explained using the primitives **contains**, **intersects**, and **overlaps-with**. We used CogSketch and VRS to compute these qualitative relations for each of the 36 images.

In evaluating whether the primitives could explain either average or individual performance, our system performed an exhaustive search for the set of weights for the primitives which maximized the Pearson correlation coefficient with the human data.

Results

Table 3 shows the correlations between the model and human ratings. As the table shows, the model correlated quite high (.98 or above) with the average human ratings for each of the four English terms. The model also correlated well with the ratings of individuals. The median correlations with individuals were all above .9. “Overlaps” was the only term for which any of the individual correlations fell below .85.

	Average	Individual Median	Individual Minimum	Individual Maximum
“Intersects”	.995	.966	.881	.999
“Overlaps”	.994	.932	.790	.999
“Connects”	.993	.95	.850	1.0
“Contains”	.981	.953	.890	.994

Table 3. Correlations between the model and human ratings of the four English terms.

There are at least two alternative explanations for the high performance of the model. One is that there is nothing special about our primitives. Perhaps any three randomly generated factors could correlate highly with human data, after performing an exhaustive search for the optimal set of weights for those factors. The other is that our model does not require all three primitives. Perhaps two of the primitives are doing all the work, and the third primitive is extraneous.

To rule out either of these possibilities, we compared our model against four other possible models (see Table 4). Three were constructed by leaving one of the three primitives out of the model and determining weights for only two primitives. The last model, Random-3 was constructed by building three random primitives (simply by randomly computing a value of *true* or *false* for each primitive’s presence in each of the 36 stimuli) and then searching for an optimal set of weights for the three random primitives. Because of the randomness involved, we constructed triplets of random primitives 40 times for each English term and averaged the results.

	C,I,O	C,I	C,O	I,O	Random-3
“Intersects”	.995	.901	.906	.994	.227
“Overlaps”	.994	.781	.956	.934	.213
“Connects”	.993	.980	.475	.993	.189
“Contains”	.981	.917	.981	.068	.217

Table 4. Several models’ correlations with the average human ratings. Letters indicate which primitives were used in each model. Random-3 uses three randomly generated factors.

As Table 4 shows, the complete model, C,I,O easily outperforms all other models. The models containing only two primitives typically perform slightly worse for three of the English terms, but each performs significantly worse on at least one term. Thus, clearly all three primitives are required to model the human rating data. The random model performs far worse than any of the other models, indicating that the particular primitives chosen for our model are much better than random factors.

Table 5 gives the optimal weights for each of the three primitives in explaining the average human ratings of the four English terms. As the table shows, there was by no means a one-to-one mapping between primitives and English terms. Participants considered at least two primitives in assessing each of the four English terms. In assessing the trickier “Overlaps” term, they appear to have considered all three primitives, on average.

	Contains	Intersects	Overlaps-with
“Intersects”	.048	.435	.516
“Overlaps”	.245	.209	.546
“Connects”	.021	.979	-.206
“Contains”	.695	0	.305

Table 5. Optimal weights of the three primitives in explaining the average human ratings of the four English terms.

Figure 4 shows the performance of the model for each individual in greater detail. In addition to showing the correlations, the figure shows the amount of weight given to each primitive by each individual. As the figure shows, there was a great deal of variation across individuals.

Discussion

As the results show, we can vary the weight assigned to each of the three primitives to create models that correlate well with either average or individual assessments of different topological terms. We can also examine the weights to see how different individuals are performing their assessments. For example, participant 7 apparently assessed “Red connects to green” entirely based on the **intersects** primitive (see Figure 4). That is, the participant believed the shapes were “connected” any time their edges intersected. On the other hand, participant 1’s model of “connects” had a strongly negative weight for **overlaps-with**. That is, the participant believed the shapes were “connected” when their edges intersected without their areas overlapping.

The results for “contains” were also quite interesting. Participant 3’s model of “contains” consists almost entirely of **contains**, indicating that the participant thought one shape contained another when the other shape was located entirely within it. However, other participants’ models of “contains” also give some weight to **overlaps-with**. This suggests that when the two circles merely overlapped, many participants believed it was somewhat appropriate to say one shape “contained” the other.

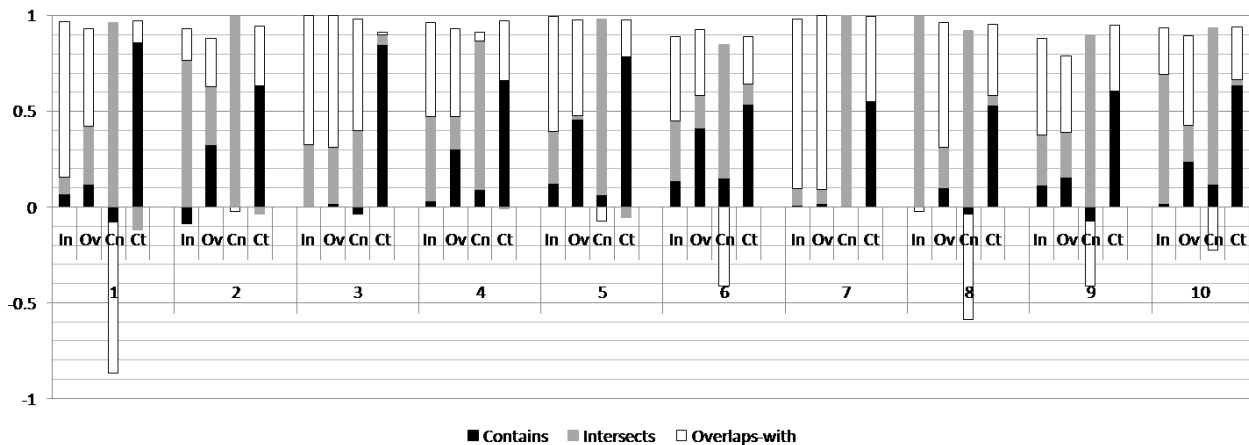


Figure 4. Model correlations with each of the 10 participants. The colors of the bars show the relative weight of each of the three primitives. In = “Intersects” Ov = “Overlaps” Cn = “Connects” Ct = “Contains”

Related Work

Lockwood and colleagues (Lockwood et al., 2006; Lockwood, Lovett, and Forbus, 2008) have used CogSketch and its predecessor along with a model of analogical generalization (Halstead and Forbus, 2005) to automatically learn representations of spatial prepositions like “on” and “in.” They have demonstrated that in both English and Dutch, the topological relation between the figure and ground plays an important role in determining which linguistic term should be used to describe the objects, although other relations like relative position are also important for some terms. They used the full set of RCC8 relations to represent topological relationships in their work.

A number of researchers have built computer models based on the idea of visual routines. However, many of these models are designed only to solve a particular problem (e.g., Chapman, 1992; Horswill, 1995), and thus they miss out on the generality promised by the original idea. Rao (1998) constructed a system for both learning and performing visual routines for solving different spatial problems. However, because his focus was on controlling a robot in the real world, the elementary operations in his system are in many cases more complex and higher-level than the simple operations proposed by Ullman.

Conclusions and Future Work

Thus far, our results support our hypothesis that individuals use three topological primitives in assessing two-dimensional topological relations. However, we have only tested the hypothesis using a small set of stimuli. In the future, we would like to expand the stimuli set to include a greater range of shapes. In particular, how do individuals assess topological relations between open shapes, e.g., lines, or between one open and one closed shape? We would also like to expand the range of English terms being assessed. However, we suspect that it will be difficult to come up with many more topological terms that can be assessed in a domain-general manner, that is, between abstract shapes. Finally, a more distant goal would be to look at how well the primitives explain topological terms from other languages.

We would also like to assess our hypothesis using a similarity rating task. In such a task, participants would see pairs of stimuli and rate their similarity on a scale from 0 to 10. Previous work has shown that people perceive stimuli as more similar or closer together when they are located in the same qualitative categories (e.g., color names: Winawer et al. 2007; or regions of a room: Newcombe and Liben 1982). Thus, by identifying similarity clusters we can better determine individuals’ qualitative categories.

One long-term question is how our two-dimensional topological primitives relate to topological relations between three-dimensional objects. We suspect that topological relations between real-world objects like those explored cross-culturally by Levinson and Meira (2003)

require integrating 2D topological primitives with both 3D depth cues and conceptual information. However, an exploration of the factors used in assessing spatial relations in three-dimensional visual scenes lies outside the scope of the present body of work.

Finally, we plan to continue developing VRS as a testbed for building cognitive models of perceptions. At present, we are concurrently evaluating a model of positional relations with VRS. Eventually, we hope to make VRS publicly available so that other researchers can use it to evaluate their own theories.

Acknowledgements

This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC). We thank our participants for volunteering to take part in our study.

References

- Chapman, D. 1992. Intermediate Vision: Architecture, Implementation, and Use. *Cognitive Science* 16: 491-537.
- Cohn, A. 1996. Calculi for Qualitative Spatial Reasoning. In J. Calmet, J. A. Campbell, and J. Pfalzgraph. eds. *Artificial Intelligence and Symbolic Mathematic Computation*, LNCS 1138, 124-143. Springer Verlag.
- Cohn, A. G., Bennet, B., Gooday, J., and Gotts, N. M. 1997. Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica* 1: 275-316.
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. 2008. CogSketch: Open-Domain Sketch Understanding for Cognitive Science Research and for Education. In Proceedings of the Fifth Eurographics Workshop on SBIM.
- Halstead, D., and Forbus, K. 2005. Transforming between Propositions and Features: Bridging the Gap. Proceedings of AAAI. Pittsburgh, PA.
- Hespos, S. J., and Spelke, E. S. 2004. Conceptual precursors to language. *Nature* 430: 453-456.
- Horswill, I. 1995. Visual Routines and Visual Search: A Real-Time Implementation and an Automata-Theoretical Analysis. In Proceedings of *IJCAI '95*.
- Kahneman, D., Treisman, A., and Gibbs, B. J. 1992. The Reviewing of Object Files: Object-Specific Integration of Information. *Cognitive Psychology* 24: 175-219.
- Kemmerer, D. 2006. The Semantics of Space: Integrating Linguistic Typology and Cognitive Neuroscience. *Neuropsychologia* 44: 1607-1621.

Landau, B., and Jackendoff, R. 2003. "What" and "Where" in Spatial Language and Spatial Cognition. *Behavioral and Brain Sciences* 16: 217-265.

Levinson, S., and Meira, S. 1993. 'Natural Concepts' in the Spatial Topological Domain—Adpositional Meanings in Crosslinguistic Perspective: An Exercise in Semantic Typology. *Language* 79: 485-516.

Li, Z. 1998. A Neural Model of Contour Integration in the Primary Visual Cortex. *Neural Computation* 10(4): 903-940.

Lockwood, K., Lovett, A., and Forbus, K. 2008. Automatic Classification of Containment and Support Spatial Relations in English and Dutch. In Proceedings of Spatial Cognition.

Lockwood, K., Forbus, K., Halstead, D., and Usher, J. 2006. Automatic Categorization of Spatial Prepositions. In Proceedings of the 28th Annual Conference of the Cognitive Science Society. Vancouver, Canada.

Newcombe, N., and Liben, L. S. 1982. Barrier Effects in the Cognitive Maps of Children and Adults. *Journal of Experimental Child Psychology* 34: 46-58.

Palmer, S., and Rock, I. 1994. Rethinking Perceptual Organization: The Role of Uniform Connectedness. *Psychonomic Bulletin and Review* 1(1): 29-55.

Pylyshyn, Z. W. 2001. Visual Indexes, Preconceptual Objects, and Situated Vision. *Cognition*, 80, 127-158.

Randell, D. A., Cui, Z., and Cohn, A. G. 1992. Naïve Topology: Modelling the Force Pump. In P. Struss and B. Faltings, eds. *Advances in Qualitative Physics*, 177-192. MIT Press.

Rao, S. 1998. Visual Routines and Attention. Ph.D. diss., Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Saund, E. 2003. Finding Perceptually Closed Paths in Sketches and Drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(4): 475-491.

Ullman, S. 1984. Visual Routines. *Cognition* 18: 97-159.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. 2007. Russian Blues Reveal Effects of Language on Color Discrimination. *Proceedings of the National Academy of Sciences* 104(19): 7780-7785.

Xu, J., and Mark, D. M. 2007. Natural Language Understanding of Spatial Relations between Linear Geographic Objects. *Spatial Cognition & Computation* 7:4, 311-347.

Yen, S. C., and Finkel, L. H. 1998. Extraction of Perceptually Salient Contours by Striate Cortical Networks. *Vision Research* 38: 719-741.