Wallgrün, J. O.; Yang, J.; Klippel, A.; Dylla, F. (in print, 2012). Investigations into the cognitive conceptualization and similarity assessment of spatial scenes. In *GIScience 2012*.

Investigations into the Cognitive Conceptualization and Similarity Assessment of Spatial Scenes[1]

Jan Oliver Wallgrün, Jinlong Yang, Alexander Klippel, Frank Dylla
Department of Geography, GeoVISTA Center, The Pennsylvania State University
[wallgrun,jinlong,klippel]@psu.edu
Cognitive Systems, Spatial Cognition SFB/TR 8, Universität Bremen
dylla@sfbtr8.uni-bremen.de

Abstract. Formally capturing spatial semantics is a challenging and still largely unsolved research endeavor. Qualitative spatial calculi such as RCC-8 and the 9-Intersection model have been employed to capture humans' commonsense understanding of spatial relations, for instance, in information retrieval approaches. The bridge between commonsense and formal semantics of spatial relations is established using similarities which are, on a qualitative level, typically formalized using the notion of *conceptual neighborhoods*. While behavioral studies have been carried out on relations between two entities, both static and dynamic, similar experimental work on complex scenes involving three or more entities is still missing. We address this gap by reporting on three experiments on the category construction of spatial scenes involving three entities in three different semantic domains. To reveal the conceptualization of complex spatial scenes, we developed a number of analysis methods. Our results show clearly that (I) categorization of relations in static scenarios is less dependent on domain semantics than in dynamically changing scenarios, that (II) RCC-5 is preferred over RCC-8, and (III) that the complexity of a scene is broken down by selecting a main reference entity.

1 Introduction

Formally capturing spatial semantics is a challenging and still largely unsolved research endeavor. Over the last two decades, a multitude of different spatial (and temporal) formalisms, often referred to as qualitative spatial calculi, have been suggested in the literature to model human commonsense understanding of spatial and spatio-temporal relations (see Cohn & Renz, 2008 for an overview). Calculi developed in the general area of qualitative spatial and temporal representation and reasoning (QSTR) allow for meaningful processing of spatio-temporal information because they focus on categorical (discrete) changes or salient discontinuities (Egenhofer & Al-Taha, 1992; Galton, 2000) in the environment, which are thought to be relevant to an information processing system (both human and artificial). While qualitative calculi are naturally appealing and, on a general level, widely acknowledged in both spatial and

[1] This research is funded by the National Science Foundation (#0924534) and Deutsche Forschungsgemeinschaft (DFG) grant SFB/TR 8 Spatial Cognition.

adfa, p. 1, 2011.

cognitive sciences (e.g. Kuhn, 2007; Lakoff & Johnson, 1980), there is comparatively little behavioral assessment of the cognitive adequacy of these calculi (see Klippel, Li, Yang, Hardisty, & Xu, in press and Mark, 1999 for overviews). To the best of our knowledge, there are no studies involving more than two entities at the same time, an observation that forms the motivation for the work described in this paper.

The two most prominent qualitative spatial formalisms in GIScience are arguably the 9-Intersection model (Egenhofer & Franzosa, 1991) and RCC-8 (Randell, Cui, & Cohn, 1992). Although, the underlying formalization is different in each approach, both formalisms make the same eight basic distinctions for topological relations holding between two simple regions in the plane (see Figure 1). When we look at applications of these and other qualitative models, many of them already employ or could benefit from incorporating a suitable notion of similarity between spatial configurations of objects. In querying and retrieval scenarios based on qualitative information (Papadias & Delis, 1997), for instance, a model of relational similarity allows for providing a ranked set of solutions (instead of returning just one solution).

The common approach to measure the similarity between two qualitative relations from the same qualitative calculus is based on so-called conceptual neighborhood graphs (CNG) (Egenhofer & Al-Taha, 1992; Freksa, 1992). CNGs are based on a notion of continuous change on a qualitative level (Galton, 2000) and two relations R_1 and R_2 are said to be conceptual neighbors if it is possible for R_1 to hold over a tuple of objects at a certain point in time, and for R_2 to hold over the tuple at a later time, with no other (third) mutually exclusive relation holding in between (Cohn, 2008). A CNG has one node for each relation and an edge between two nodes if the corresponding relations are conceptual neighbors. In Figure 1, the edges show the CNG structure of RCC-8 and the 9-Intersection model.

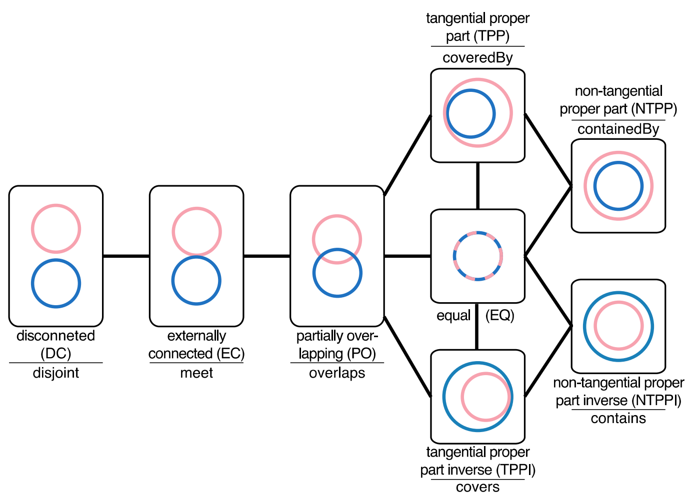

Figure 1: Relations of RCC-8 and the 9-Intersection calculus arranged in accordance with their conceptual neighborhood graph (indicated by the edges).

Traditionally, the dissimilarity (or distance) $d(R_1, R_2)$ between two relations has been measured by assuming uniform weights for the edges in the CNG and counting the number of elementary changes or steps along the shortest connecting path in the CNG (Bruns & Egenhofer, 1996; Schwering, 2007). The dissimilarity of RCC-8 relations DC and PO, for instance, is 2 while it is 4 for DC and NTPP (see Figure 1). However, this simplistic approach has been challenged: On the one hand, researchers have developed alternative approaches using different weighting schemes, mainly based on intuition and introspection such as in the work by Li and Fonseca (2006): a weight of 3 is assigned to the edge between DC and EC, a weight of 2 for EC and PO, and a weight of 1 for TPP and NTPP. Only a few empirical investigations on the appropriateness of qualitative calculi using, for instance, grouping experiments with visual stimuli (Mark & Egenhofer, 1994) have been undertaken with the goal of painting a clearer picture of human relational similarity assessments and its relation to qualitative spatial formalisms. Related to this is the question whether the relational equivalence classes introduced by a qualitative calculus make the relevant distinctions to begin with or whether, for instance, coarser models such as RCC-5 or the coarse version of the 9-Intersection model (Knauff, Rauh, & Renz, 1997) should be preferred.

While progress has been made over the last years in evaluating the appropriateness of qualitative calculi and grounding similarity weighting of the respective relations in empirical data, we are facing a lack of similar work with respect to the problem of defining suitable similarity measures for complex spatial scenes. *Complex* is defined here as spatial configurations involving more than two objects. This fact is astonishing as such measures are urgently needed for application areas such as similarity-based querying and retrieval. Existing computational approaches (Bruns & Egenhofer, 1996; Dylla & Wallgrün, 2007; Papadias & Delis, 1997) compute similarities between qualitative equivalence classes (QECs) defined by the $m = n(n-1)/2$ qualitative relations holding between *n* spatial entities by aggregating, in particular summing up, elementary neighborhood distances over corresponding relations, for example:

$$D(QEC_1, QEC_2) = \sum_{i=1}^{m} d(R_i^{[1]}, R_i^{[2]})$$

where $R_i^{[1]}$ and $R_i^{[2]}$ stand for the ith relation from QEC_1 and QEC_2, respectively. With eight base relations in RCC-8, there exist 512 possible equivalence classes for three entities; but, only 193 of these are consistent QECs in the sense that they can be satisfied by actual triples of simple regions in the plane. Figure 2 shows approximately 15% of these 193 QECs depicted by an exemplary configuration of three ellipses with the respective qualitative relations listed on the side. The QECs for *n* entities can be connected to form a conceptual neighborhood graph (called CCNG for complex conceptual neighborhood graph) in the same way as the CNG for individual relations. The edges in the depicted CCNG connect those QECs in which exactly one of the relations has changed to a conceptual neighbor (e.g., EC to PO). The connected pairs of QECs are exactly those for which the dissimilarity $D(QEC_1, QEC_2)$ is 1. This, however, raises many important questions with regard to a suitable choice for the involved aggregation operators as well as the appropriateness of the overall approach.

An empirical basis to answer these questions, which can be expected to improve current implementations, is still largely missing.

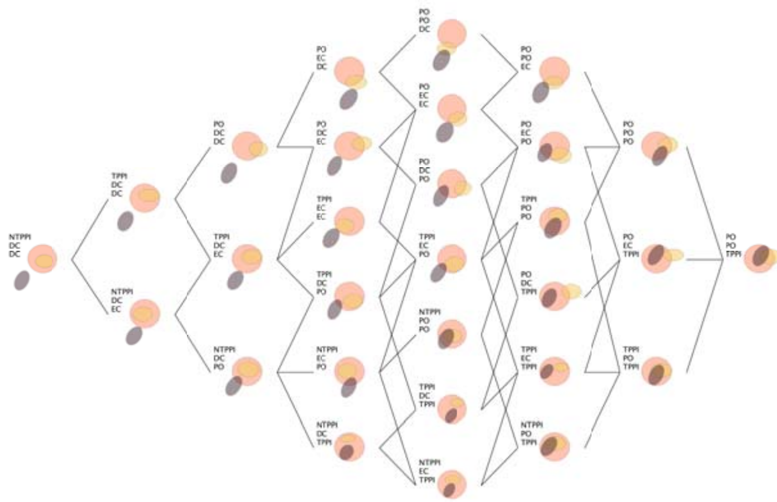

Figure 2. Part of the RCC-8 / 9-Intersection complex neighborhood graph for three objects. Edges connect those QECs that have an aggregated distance of $D = 1$

The research described in this paper aims at remedying this situation by developing the empirical and methodological basis for evaluating and improving qualitative approaches for relational similarity assessments in complex scenes involving more than two objects. We report on three grouping experiments in which participants were given icons showing different configurations of three simple objects (Section 2). In our analysis (Section 3), we employ different clustering and cluster validation approaches to compare human similarity assessment (as an expression of cognitive conceptualizations) to the qualitative equivalence classes induced by topological calculi and evaluate the adequateness of the approach using $D(QEC_1, QEC_2)$ as defined above as a model of similarity.

2 Experiments

This section details three category construction (grouping) experiments that we conducted to shed light on human conceptualizations of spatial scenes with three entities. For the purpose of this paper, we define a spatial scene as a configuration of three spatially extended objects visually represented in a map-like format. Specifically, we used three elliptical entities such that each scene can be characterized using three topological relations (see Figure 2). Each of the three experiments was identical except for the semantic domain information that was associated with the scenes. The semantic domains we chose for this experiment in addition to a purely geometric one were: an ocean scenario with a blue water background and a forest scenario with a greenish background. The ellipses themselves were introduced as areas demarcating habitats of species. Figure 3 shows two instances of icons for each of the three domains (geometric, ocean, forest).

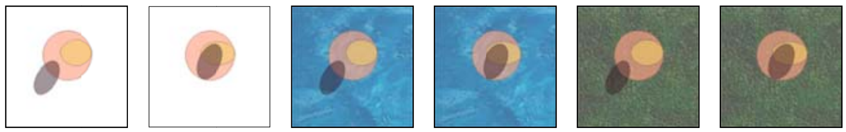

Figure 3. Two exemplary icons for each of the domains: geometric, ocean, and forest.

Design and materials. The visual stimuli used in the experiments consist of three sets of 116 icons each, one set for each of the three semantic domains (see Figure 3). In our experiments, we consider the 29 QECs shown in the partial CCNG from Figure 2. These are all QECs that can be considered as being 'between' the QEC NTPPI-DC-DC on the left and the QEC PO-PO-TPPI on the right. For each of these 29 QECs we created four instances that were topologically identical but varied geometrically; hence, 4 x 29 = 116 icons. Each icon was 120x120 pixels in size. The geometric layout was randomized in the following way: We started with the geometric configurations representing the respective QEC in Figure 2 and randomized the following parameters of the two smaller ellipses: semi-major radius, semi-minor radius, *x* and *y* coordinate of center, and rotation angle. This was done using a uniform probability distribution over the interval $[p - \delta, p + \delta]$ where p is the value of the parameter in the prototype and delta is an individually chosen threshold value. The threshold values used were 10 pixels for the coordinates, 4 pixels for both radii, and 5 degree for the rotation angle. Because the random variation may change the qualitative relations holding between the objects, this step was followed by a brute force search within the parameter space for a set of parameters closest to the randomly generated parameter set and satisfying the qualitative relations given in the respective QEC. Possible parameter sets were constrained by the fact that some topological relations (e.g., NTPPI) are only possible for certain size relations between the involved entities.

We originally generated 10 geometric instances for each QEC and generated icon sets by drawing the ellipses on different backgrounds (white background for the purely geometric domain, and textured backgrounds for the other two domains) using transparency and a color scheme that would work well with all three different backgrounds. We then manually selected the first four icon instances for each QEC that were visually clear in the sense that the qualitative relations were deemed to be recognizable. It turned out that for three QECs additional instances had to be generated to get four clearly recognizable scenes.

Participants. Each experiment had 22 participants, Penn State students who received course credit for their participation. The female-to-male ratios were 11/11 for the geometry condition, 8/14 for forest, and 11/11 for ocean (one participant had to be excluded because of providing bizarre information in the linguistic descriptions, see Procedure). The average age was 22.75, 20.05 and 21.05, respectively.

Procedure. The experiments were designed as group experiments and took place in a GIS lab. Up to 16 participants were able to take part in the experiments at the same time with workplaces separated by view blocks. Computers were Dell workstations with 24” widescreen LCD displays. The experiment was administered through our custom made software CatScan (Klippel, Li, Hardisty, & Weaver, 2010). Participants only grouped one of the three scenarios and were explicitly introduced to the semantics of the scenario that they were supposed to imagine. To ensure that they understood the task and semantics of the scenario, they had to enter keywords (e.g., forest, habitat) into the interface before they could start the experiment. Keywords were checked for their correctness. They also were given an unrelated category construction task (Medin, Wattenmaker, & Hampson, 1987) to acquaint themselves with the general idea of category construction and the interface. Participants then performed the category construction task on the stimuli. All 116 icons were initially presented on the left side of the screen with no groups on the right side. Participants were required to create all groups (as many as they thought appropriate) themselves. CatScan allows for icons to be moved around (into, out off, and between groups) by a simple mouse drag and drop procedure. After sorting all icons into group(s), participants were again shown the groups they had created and asked to provide a short linguistic label (max. 5 words) and a more detailed description of their grouping rational.

3 Results

The data we collected in the three experiments comprised information about the categories each participant created in the form of binary matrices ranging over the icon sets containing a ‘1’ if the respective icons were put into the same group and a ‘0’, otherwise. These matrices form the basis for the analyses conducted and described in this section. In addition, the linguistic descriptions were collected in spreadsheets.

Our analysis and evaluation described in this section addresses the question of the influence of domain semantics as well as a detailed analysis of the category construction behavior of participants. The latter can be taken as a basis for evaluating existing approaches on defining similarities (semantics) of spatial scenes.

3.1 Comparison of Raw Similarities

To derive overall raw similarities for each of the three experiments, we combine the binary matrices from individual participants into a single overall similarity matrix (OSM) by summing up corresponding matrix cells. As a result, we get a matrix with values from 0 for pairs of icons that were never put into the same group (and, hence, are rated as maximally dissimilar) to N (= number of participants; here: 22) for pairs that were put into the same group by all participants, considered to be maximally similar. Figure 4 illustrates the resulting OSMs in form of heat maps using colors from white (corresponding to 0) to red (corresponding to N). The entries are alphabetically ordered such that all 4 icons belonging to the same QEC correspond to a group of neighbored rows and columns in the matrices.

The heat maps allow for a first visual inspection of the grouping behavior. The red 4x4 squares along the diagonals of all three matrices are a clear indication that icons belonging to the same QEC are rated as being very similar, that is, they are (almost) always placed together into the same group. To back up this observation by numbers, we computed the sum over all entries for each block (QEC), took the average over all QECs, and normalized the result to be within [0,1]. The results show an average of 0.95 with 0.03 standard deviation for the purely geometric domain, 0.93 for ocean (standard deviation 0.04), and 0.94 for forest (standard deviation 0.04). This can be interpreted as evidence that topological equivalence classes potentially offer an explanation of how humans conceptualize spatial scenes. Additionally, however, there are several other areas in the OSMs with high similarities. This is a first indication that the 29 topologically defined equivalence classes form coarser conceptual groups.

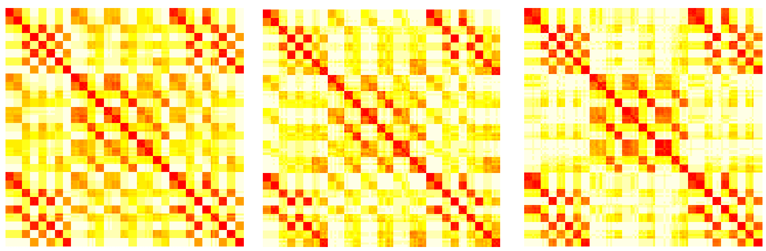

Figure 4. Heat maps showing raw similarities (red = maximally similar, white = maximally dissimilar) for geometry (left), ocean (middle) and forest (right).

Further comparison of the heat maps in Figure 4 shows that overall the three patterns are very similar. To make, however, the differences more explicit, we computed difference-matrices for each pair of OSMs using the operation abs(OSM_1-OSM_2) for each cell. The resulting matrices are shown in Figure 5 emphasizing where differences do exist. Computing the average differences over all entries (except the diagonals which have to be zero) and normalizing them to [0,1], we get the following results: 0.08 for geometry-ocean, 0.08 for geo-forest, and 0.09 for forest-ocean. This means that the difference in similarity assessment averaged over all pairs of icons is less than 9% between the domains. This is a very low number given that within each domain individual differences exist, too.

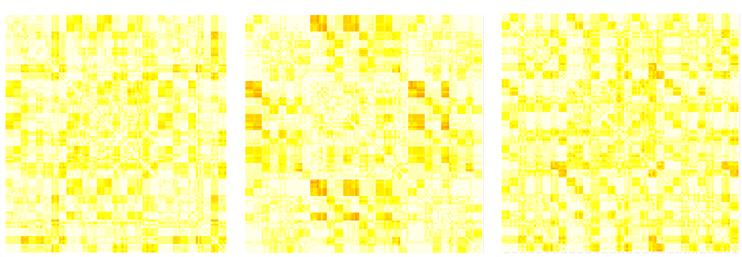

Figure 5. Heat maps showing the differences between OSM matrices for geometry-ocean, geometry-forest, and forest-ocean (white = 0 difference; maximal difference would be red but does not occur).

3.2 Clustering

We followed widely accepted procedures on cluster analysis and cluster validation, that is, for each scenario we performed three different types of cluster analysis (Ward's methods, average linkage, complete linkage) and compared the clustering structure (Kos & Psenicka, 2000). The resulting clusterings can be visualized as tree structures called dendrograms in which the leaf nodes represent the individual icons (instances of a QEC). Figure 6 shows a small part of such a dendrogram. We found large similarities between the different scenarios but also dissimilarities especially comparing different methods. The reasons for these differences seem to be largely unrelated to the semantics of a particular domain but are the results of a more complex decision space: We have 29 QECs with four instances for each QEC. Looking into how hierarchical cluster algorithms operate, we find that initial similarities/dissimilarities can lead to different clustering structures reinforced by the recalculation of similarities after each clustering step; these differences are not reflective of high overall similarities (as Figure 5 shows that all scenarios are very similar). Given space constraints, it is not possible to discuss all nine cluster analyses (three for each experiment/domain) in detail. However, to harvest what cluster analysis reveals about the similarities / categories of complex spatial scenes, we developed a method that we consider highly valuable for researchers evaluating results of clustering methods. With this method that we term *greatest common divisor* algorithm, we are able to identify the most fundamental category construction aspects (similarities) across all nine cluster analyses (three scenarios with three cluster analyses each).

An important prerequisite is that QECs are very strong predictors for category construction, that is, instances of a QEC are not separated in any of the clustering methods (compare Section 3.1). Either all icons of the same QEC are combined into a single cluster before the resulting cluster is combined with icons from a different QEC; or, icons from two or, in a few cases, three conceptually neighbored QECs are joined in a merged way forming a single cluster. This is another indication that topological relations and conceptual neighborhood graphs capture important factors of the cognitive conceptualization of spatial scenes.

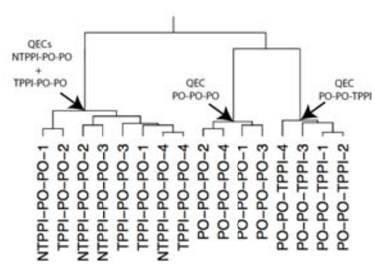

Figure 6. Part of a dendrogram from a clustering method. The leaf nodes represent the icons (instances of QECs) which are combined to form larger groups on higher levels.

1: **procedure** GREATESTCOMMONDIVISOR($\mathcal{D}$)

2: **Input:**
$\mathcal{D}$ set of dendrograms D_i with leaf nodes annotated with $\{s\}$ where s is the respective icon name

;; initialization

3: **do**

4: find a node N on level depth$(D_i) - 1$ in a $D_i \in \mathcal{D}$ and with either
not all child nodes of N be labeled with the union of all icons of a set of QECs
or the label of a child node of N is a strict subset of the label of a leaf node in another D_j

5: call MERGE(D_i, N)

6: **until** no such node can be found

;; clustering

7: **do**

8: find one N_i for each $D_i \in \mathcal{D}$, all on level depth$(D_i) - 1$ and with identically labeled child nodes

9: call MERGE(D_i, N_i) for all i

10: **until** no such set of nodes can be found

11: **end procedure**

1: **procedure** MERGE(D, N)

2: **Input:**
D tree
N node in D at level depth$(D) - 1$

3: annotate N with the union of annotations of its child nodes

4: remove all child nodes of N

5: **end procedure**

Algorithm 1. Algorithm to derive largest common clusters.

Now that we know that individual QECs are potential category predictors, we seek to find QECs most similar to each other. To this end, we continued the bottom up analysis of consistent clustering results across all three scenarios and all clustering methods using the greatest common divisor algorithm shown in Alg. 1. This algorithm aims at determining the largest groups of QECs for which the order of combination is identical over all three experiments and all three clustering methods. It consists of two phases: the initialization and the main loop (clustering).

In the initialization phase, the algorithm merges leaf nodes starting with the individual icons, until we have nine tree structures with identical leaves in terms of associated icons, and each leaf represents all icons from one or more QECs. In Figure 6, for instance, we end up with the new leaf nodes marked by the arrows with the left one representing two QECs and the other two representing individual QECs.

In the main loop, the algorithm combines leaf nodes if they exist in all trees and are connected to the same parent node, meaning that clusters are merged in the same local order along rising branches in all dendrograms. The result of this procedure applied to our nine dendrograms is shown in Fig. 7 (colored lines surrounding QECs). Several larger groups are identified using this algorithm that are plausible from the perspective of applying a coarser calculus such as RCC-5 (we will come back to this discussion in the conclusion/discussion section). However, we also find that several parts of the CCNG are split up to form clusters with only one or two QECs. To address this issue, we modified Alg. 1 to use a less restrictive criterion in line 8: We allowed merging when the respective groupings were combined in at least six out of nine cluster analyses. Using this modification, we were able to reveal some essential aspects of the grouping behavior of participants across all three experiments.

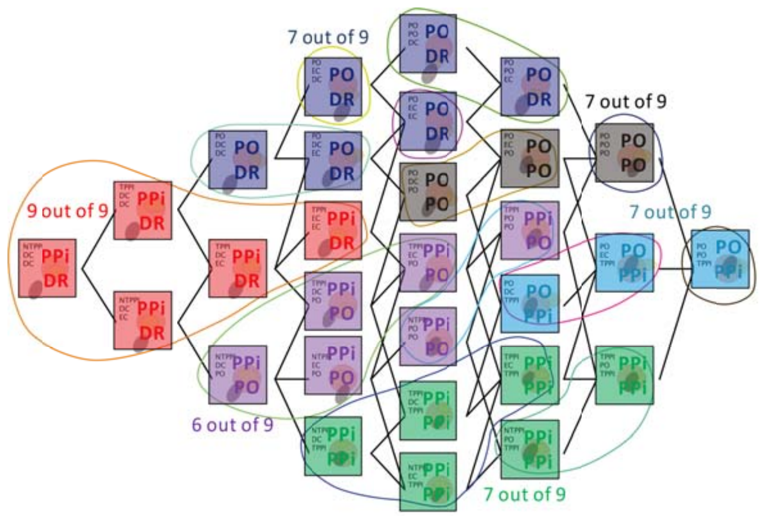

Figure 7. Greatest common divisor clusters identified by Algorithm 1. Colored lines indicate results using hard constraints, that is, clusters found all 9 dendrograms. Colored boxed indicate relaxed constraints (minimum 6 out of 9 dendrograms). Colored letters (e.g., DR) indicate RCC-5 relations (not for relations between small ellipses).

To further visualize the results, we annotated Figure 7 in the following way: 1) we used a color coding to identify the six resulting cluster; 2) each scene has been characterized based on a coarser level of granularity (bold letters) using a) only the relation between the large circle and each of the smaller ellipses (but not the relations between the two smaller ellipses); b) instead of using RCC-8, we used RCC-5 which combines DC and EC into DR (discrete) and NTPP and TPP to PP (proper part).

With these annotations and reanalysis we were able to reveal an astonishingly clear picture of the grouping behavior of participants: The conceptualization process of participants centered on two aspects: First, relations were simplified according to RCC-5, second, the important distinctions were made looking into the relationship between each small ellipse with the larger circle individually while the relation between the smaller ellipses was largely ignored. This strategy was not scenario specific as in most cases seven (or more) cluster methods support this interpretation.

3.3 Linguistic Analysis

We performed a word-count analysis on the linguistic descriptions collected using AntConc, a corpus analysis toolkit (Anthony, 2011). Before the analysis, we excluded spatially irrelevant words such as colors (e.g., red, black, yellow), pronouns (e.g, this, these, which), words referring to the entities (e.g., ellipse, oval, habitat), and other common English words (e.g., is, but, and). In addition, we combined the frequencies of the same word in different tenses (e.g., overlap, overlaps, overlapping, and overlapped), and also synonyms (e.g., completely, fully, totally, and entirely). The final results are shown in Table 1.

Table 1. Top 10 frequently mentioned words from participants' linguistic description.

	Geometry		**Forest**		**Ocean**	
Rank	*Word*	*Frequency*	*Word*	*Frequency*	*Word*	*Frequency*
1	in	130	overlap	116	overlap	102
2	overlap	97	in	114	in	91
3	partially	90	completely	96	both/two	55
4	both/two	89	partially	80	inside	54
5	completely	89	both/two	76	completely	49
6	inside	86	inside	73	within	49
7	out	69	with	51	outside	46
8	touching	58	not	47	all	45
9	outside	53	outside	47	not	44
10	not	41	all	30	with	37

First, it is noteworthy that the words "in", "inside", "out", and "outside" are most frequently mentioned across the three domains. This suggests that the non-overlapping relation (DC and EC) are distinguished from overlapping relations (TPPI and NTPPI). Second, the only word referring to connecting relations (EC and TPPI) is "touching" (ranked 9th in geometry scene), which may indicate that the connecting relation is more relevant in the geometric domain (compared to forest and ocean). Third, "both" and "two" are frequently used by participants across all semantic domains. By additionally looking into the original descriptions, we found that, in most cases, these two words are used to describe the relations of the two smaller entities to the larger entity in each scene. The abovementioned findings support the conclusions we drew from the cluster analysis, i.e., participants' overall grouping rationale relies on RCC-5 and the relation between two smaller entities is often ignored.

4 Discussion and Conclusions

Constructing categories is arguably one of the most fundamental abilities that humans possess. Paralleling this aspect, the disciplines of the spatial sciences focus strongly on conceptualization and categorization to structure spatial as well as temporal information, often using ontological frameworks (Bateman, Hois, Ross, & Tenbrink, 2010). In the spatial sciences and related branches of artificial intelligence, qualitative spatio-temporal representation and reasoning formalisms play a prominent role in connecting human category construction with formal approaches to advance processes at the human-machine interface (representation, reasoning, retrieval).

The research reported in this paper closes an important gap: While approaches on simple configurations exist, no data is available on more complex scenarios, here: relations between three entities. Explorations into more complex and real world scenarios are important: First, because discontinuities identified by qualitative calculi focusing on two relations may behave differently in complex scenes with more relations (e.g., similarities/dissimilarities may or may not be adding up directly); and second, because it is not clear whether and how domain semantics influence the conceptualization of static spatial relations (see Coventry & Garrod, 2004 for a general discussion and Klippel, accepted for dynamic processes).

The results reported here can be summarized as follows: 1) Topological equivalence is a strong grouping criterion / category predictor. This is prominently demonstrated by the analysis of the grouping behavior of instances within QECs that are almost always placed together into the same groups. 2) Overall, the similarities between all three scenarios are highly indicating that—in this static case—the semantics of individual domains may not play a substantial role on the construction of categories of spatial relations, at least not for the domains chosen here. This analysis is reinforced by the linguistic descriptions provided by participants. They reveal that participants placed a strong focus on purely spatial aspects rather than incorporating domain specific language (other than referring to ellipses by using their color). 3) As the decision space gets more complex in CCNGs, there is more variation across different experiments and classic clustering methods are not necessarily well suited to distinguish commonalities from differences. To address this issue, we designed an algorithm that revealed the most fundamental coarse categories constructed by participants by comparing (here) nine different cluster analyses (three for each experiment/domain). We were able to demonstrate, clearly, two factors that explain the category construction behavior of participants: RCC-5 works well as a predictor of category membership taking additionally into account that the largest entity was used as a reference. As a result, the relations between the two smaller ellipses only played a subordinate role. We found the clarity of these results quite surprising. 4) Within all experiments and all cluster analyses (the original nine, not the aggregated one), we did never find a violation of category membership induced by the CCNG. In other words, all members of groups identified in the nine cluster analyses are always neighbors in the CCNG. This is probably one of the most promising results as it adds to the validity of using CCNGs for similarity assessments and category prediction.

These results support existing theories on conceptualization and category construction for spatial and non-spatial information. It has been a long debate how humans deal with complexity (Heil & Jansen-Osmann, 2008). Across different disciplines it is generally assumed (and experimentally confirmed) that humans will reduce complexity and lower the individual pieces of information that they have to deal with (Cowan, 2001). In the end, this is what categorization is all about. What is less clear is which mechanisms they use and how to formally describe them such that they may be used in artificial systems, too. Two approaches are worth considering: a) participants could try to holistically assess the similarity of the scenes we presented them with; b) participants single out a particular dimension along which they construct categories (Pothos & Close, 2008). While both approaches are mutually exclusively discussed in the literature, our results seem to indicate that participants used a combination of both strategies. On the one hand, they singled out aspects (dimension in a looser interpretation) that they were able to use as anchors to categorize the scenes, specifically, a reduction of three relations to two by ignoring the relations between the smaller ellipses. On the other hand, they holistically simplified the scenes by ignoring RCC-8 and adopting a coarser perspective that can be captured by RCC-5.

Based on the promising results we will pursue this line of research to assess spatial similarity on different levels of scene complexity to advance approaches to formalize spatial semantics. We will perform additional experiments with, for example, varying domains and relaxation of the spatial constraints which we applied in the current experiments (e.g., to include additional aspects of spatial knowledge). One critical topic will be to investigate how the similarity measures derived from behavioral data can be transformed best into weights in (complex) conceptual neighborhood graphs.

References

Anthony, L. (2011). *AntConc (version 3.2.2)*. Tokyo, Japan: Waseda University, available from http://www.antlab.sci.waseda.ac.jp/.

Bateman, J. A., Hois, J., Ross, R., & Tenbrink, T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence, 174*(14), 1027–1071.

Bruns, H. T., & Egenhofer, M. J. (1996). Similarity of spatial scenes. In M. J. Kraak & M. Molenaar (Eds.), *Seventh international symposium on spatial data handling (SDH'96), Delft, The Netherlands* (pp. 173–184).

Cohn, A. G. (2008). Conceptual neighborhood. In S. Shekhar & H. Xiong (Eds.), *Encyclopedia of GIS* (p. 123). Boston, MA: Springer.

Cohn, A. G., & Renz, J. (2008). Qualitative spatial representation and reasoning. In F. van Harmelen, V. Lifschitz, & B. Porter (Eds.), *Foundations of artificial intelligence. Handbook of knowledge representation* (1st ed., pp. 551–596). Elsevier.

Coventry, K. R., & Garrod, S. (2004). Towards a classification of extra-geometric influences on the comprehension of spatial prepositions. In L. A. Carlson & E. van der Zee (Eds.), *Functional features in language and space.* Oxford UP.

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences, 24*, 87–186.

Dylla, F., & Wallgrün, J. O. (2007). Qualitative spatial reasoning with conceptual neighborhoods for agent control. *Journal of Intelligent and Robotic Systems, 48*(1), 55–78.

Egenhofer, M. J., & Al-Taha, K. K. (1992). Reasoning about gradual changes of topological relationships. In A. U. Frank, I. Campari, & U. Formentini (Eds.), *Theories and methods of spatio-temporal reasoning in geographic space.* Springer.

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems, 5*(2), 161–174.

Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence, 54*(1), 199–227.

Galton, A. (2000). *Qualitative spatial change. Spatial information systems.* Oxford: Oxford Univ. Press.

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *The Quarterly Journal of Experimental Psychology, 61*(5).

Klippel, A. (accepted). Spatial information theory meets spatial thinking - Is topology the Rosetta Stone of spatio-temporal cognition? *Annals of the Association of American Geographers*, (67 manuscript pages).

Klippel, A., Li, R., Hardisty, F., & Weaver, C. (2010). Cognitive invariants of geographic event conceptualization: What matters and what refines. In S. I. Fabrikant, T. Reichenbacher, M. van Krefeld, & C. Schlieder (Eds.), *Sixth International Conference, GIScience, Proceedings* (pp. 130–144). Berlin: Springer.

Klippel, A., Li, R., Yang, J., Hardisty, F., & Xu, S. (in press). The Egenhofer-Cohn Hypothesis: Or, Topological Relativity? In M. Raubal, A. U. Frank, & D. M. Mark (Eds.), *Cognitive and Linguistic Aspects of Geographic Space - New Perspectives on Geographic Information Research*.

Knauff, M., Rauh, R., & Renz, J. (1997). A cognitive assessment of topological spatial relations: Results from an empirical investigation. In S. C. Hirtle & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS.* Springer.

Kos, A. J., & Psenicka, C. (2000). Measuring cluster similarity across methods. *Psychological Reports, 86,* 858–862.

Kuhn, W. (2007). An image-schematic account of spatial categories. In S. Winter, B. Kuipers, M. Duckham, & L. Kulik (Eds.), *Spatial Information Theory. 9th International Conference. Proceedings* (pp. 152–168). Berlin: Springer.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Li, B., & Fonseca, F. (2006). TDD: A comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation*, *6*(1), 31–62.

Mark, D. M. (1999). Spatial representation: A cognitive view. In D. J. Maguire, M. F. Goodchild, D. W. Rhind, & P. A. Longley (Eds.), *Geographical information systems: Principles and applications. Second edition, v. 1* (pp. 81–89).

Mark, D. M., & Egenhofer, M. J. (1994). Calibrating the meanings of spatial predicates from natural language: Line-region relations. In T. C. Waugh & R. G. Healey (Eds.), *Advances in GIS Research.* (pp. 538–553).

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*(2).

Papadias, D., & Delis, V. (1997). Relation-based similarity. In *Proceedings of the 5th ACM Workshop on GIS, Las Vegas, ACM* (pp. 1–4).

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition*, (2), 581–602.

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connections. In B. Nebel, C. Rich, & W. R. Swartout (Eds.), *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning* (pp. 165–176). San Francisco: Morgan Kaufmann.

Schwering, A. (2007). Semantic similarity of natural language spatial relations. In *Conference on Artificial Intelligence and Simulation of Behaviour: Artificial and Ambient Intelligence. Symposium: Spatial Reasoning and Communication.*