Investigating Intuitive Granularities of Overlap Relations

Jan Oliver Wallgrün, Jinlong Yang, Alexander Klippel
Department of Geography, GeoVISTA Center
The Pennsylvania State University
University Park, PA 16802
Email: {wallgrun, jinlong, klippel}@psu.edu

***Abstract*—We present four human behavioral experiments to address the question of intuitive granularities in fundamental spatial relations as they can be found in formal spatial calculi that focus on invariant characteristics under certain (especially topological) transformations. Of particular interest to this article is the concept of two spatially extended entities overlapping each other. The overlap concept has been extensively treated in Galton's mode of overlap calculus [1]. In the first two experiments, we used a category construction task to calibrate this calculus against behavioral data and found that participants adopted a very coarse view on the concept of overlap, only distinguishing between three general relations: proper part, overlap, and non-overlap. In the following two experiments, we changed the instructions to explicitly address the possibility that humans could be swayed to adopt a more detailed level of granularity, that is, we encouraged them to create as many meaningful groups as possible. The results show that the three relations identified earlier (overlap, non-overlap, and proper part) are very robust and a natural level of granularity across all four experiments but that contextual factors gain more influence at finer levels of granularity.**

I. Introduction

Research on granularities has a long tradition in cognitive computing and cognitive informatics. Tasks that require to flexibly change between different levels of granularity are omnipresent in the cognitive organization of spatial knowledge and are reflected in numerous computational models and applications [2], [3], [4].

An important research endeavor in this context is the identification of intuitive levels of granularity in human conceptualizations and commonsense understanding of spatial relations. The understanding of this key part of human spatial cognition plays a crucial role in the design of intuitive user interfaces and of intelligent spatial reasoning and assistance systems. We can find different levels of granularity in several formal models of spatial relations, so-called spatial calculi, such as the prominent topological Region Connection Calculus (RCC) [5] and Intersection Model (IM) [6] calculus families. The same holds true for formalisms dealing with direction or orientation information where we can observe a recent trend towards approaches whose granularity is adaptable (see [7], [8]). In addition to capturing human intuition and categorization behavior, these approaches aim at improving the computational efficiency of processing spatial and temporal information by providing an adequate level of granularity sufficient to solve a particular task or for performing a task in a hierarchical manner. Overall, granularity is a key element of processing information or communicating it to a human user where involving too much detail would violate principles of cognitive ergonomics as discussed, for example in Clark's 007 Principle: "In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done." ([9], p. 64).

While spatial calculi are widely acknowledged in both spatial and cognitive sciences (e.g., [10], [11]), there is comparatively little behavioral assessment of the question whether the distinctions made in the suggested formalisms, hereafter referred to as qualitative equivalence classes (QEC), actually are suitable to capture human intuition and categorization (see [12], [13] for overviews). In our research, we aim at remedying this situation by comparing the QEC-implied categories for fundamental spatial concepts with the cognitive behavioral data collected in free classification experiments in order to prove, disprove, or calibrate the cognitive adequacy of a formal model. In this article, we are addressing the question of intuitive granularities for the fundamental spatial concept of *modes of overlap* as formalized in a model by Galton [1] employing crowdsourcing approaches. To this end, we have designed several experiments that challenge findings that *overlap* is a largely undifferentiated concept from an intuitive or commonsense perspective. In the following, we will briefly discuss the notion of fundamental spatial concepts and the role of crowdsourcing in conducting experiments important for cognitive informatics. We continue by giving a short introduction to the *modes of overlap* calculus by Galton [1] and move on to detail the four experiments. We proceed with an analysis and presentation of the main results of our experiments, followed by a discussion and conclusions.

A. Fundamental Spatial Concepts

The notion of "fundamental spatial concepts" has been frequently discussed in the literature (see, for instance, [14], [15]). While there is no commonly accepted definition, the concept of *invariants* wrt. certain kinds of mathematical transformation provides a means to render it more precisely. The topic of invariants and their application for information processing and for defining formal distinctions between spatial concepts has been discussed by researchers from different fields, such as

This research is funded through the National Science Foundations under grant number #0924534.

cognitive and spatial sciences. Klein introduced the distinction between different geometries based on properties that remain invariant under particular transformations (see Erlangen program [16]). This allows for a distinction between Euclidean geometry, set-based geometry, and topology where the latter is concerned with properties that are preserved under topological transformations such as bending and stretching. This approach has been adopted by Worboys and Duckham as a framework for their chapter on Fundamental Spatial Concepts in [14]. The cognitive science community has always been interested in perceptual and cognitive invariants, which we also find to be associated with conceptual primitives (e.g., [17]). For instance, Klix [18] states that the human mind identifies invariant characteristics of information as the basis for cognitive processes when adapting to its environment. Another example is the work by Shaw [19] who uses the term transformational invariants to denote properties of objects and events that do not change from a group (set) theoretical perspective, stressing the importance of this concept for perception. Gibson, in his seminal work [20], refers to temporally constant characteristics of environments as structural invariants. His work has influenced many recent approaches utilizing the concept of invariants to explain, for instance, categorization and event cognition [21], [22]. Galton [23] also notes the importance of identifying invariants of environmental information and points out human's ability to intersubjectively identify invariants of space and time which allows us to construct a shared understanding of our physical (and social) environments. Without this agreement on certain characteristics of spatial environments that ground our meaningful understanding of spatial environments [24], the concept of a shared reality and our ability to communicate about this reality would not be possible. In the previously mentioned work by Klix [18], topology is proposed as one of the best candidates for identifying invariants, a fact that is reflected by the role that topological aspects play in formal modeling approaches (see, for instance, [25], [10], [26], [27]).

B. *Crowdsourcing for Behavioral Studies*

The advent of crowdsourcing technology and platforms (e.g., Survery Monkey, Lime Survey, and Amazon's Mechanical Turk (AMT)) over the last decade brought about the potential to change and substantially add to the way that knowledge about human spatial and spatio-temporal conceptualizations and representations can be investigated empirically [28]. In particular, AMT has developed into a serious alternative to using laboratory experiments and, as a result, received a lot of attention from researchers involved in behavioral research (e.g., [29], [30]). In AMT, simple tasks referred to as HITs (Human Intelligence Tasks) are outsourced to a workforce of registered users (often called "Turkers") who anonymously perform the task and return the results to AMT for review by the requester. Additionally, restrictions can be placed by allowing only Turkers above a certain HIT approval rate, obtained through successfully completing experiments. If the work by a Turker is accepted by the requester, he/she will receive payment directly through AMT. Additional bonus payments for high quality work are also an option.

An important question regarding the utilization of AMT for behavioral experiments is how comparable results are to regular laboratory experiments. While several studies (e.g., [31]) indicate that this is indeed the case, our own pre-studies suggest that some caution is advisable. Our conclusion from several test experiments is that special measures and strategies have to be employed to ensure a high quality of the obtained experimental data. In particular, one has to put oneself in the position of the Turkers which typically is to maximize financial gain over time by finding the fastest yet correct way to perform a HIT. As a result, potential strategies of Turkers need to be taken into account and channeled when designing a behavioral study for AMT and setting up the reward system.

C. *The Modes of Overlap Calculus*

In this paper, we focus on topological relationships and the fundamental concept of modes of overlap between spatially extended regions as defined in the approach of Galton [1]. This approach has been developed in response to the somewhat indifferent or unspecific treatment of the concept of overlap in the most prominent topological spatial calculi, the already mentioned Region Connection Calculus, RCC [5], and Egenhofer's Intersection Models, IM [6]. The RCC-8 and 9-IM variants of these calculi distinguish the same eight elementary spatial relations. Some of these relations are present in the table shown in Figure 2, namely as relations "a", "aa", "c", "d", "dd" and "g" (for a detailed treatment see, for instance, [32]). These theoretical topology-based distinctions have been extensively analyzed and investigated in behavioral studies [33], [34], [35]. Galton's approach can be seen as an extension of Egenhofer's intersection matrix that allows for a more differentiated categorization of topological relations. For instance, it allows for distinguishing between the relations "g" and "q" in Figure 2 where the *number* of disjoint components of the intersection of the two involved regions varies, while both are subsumed by the same relation in RCC and IM based on the fact that there *exists* an overlap between the two regions. Galton's modes of overlap relations are based on a 2×2 matrix that essentially counts the numbers of regions with different overlap conditions between the two involved entities. More precisely, the matrix is defined as

$$[A, B] = \left(\begin{smallmatrix} x & a \\ b & o \end{smallmatrix}\right)$$

where

x is the number of connected components of the intersection $A \cap B$

a is the number of connected components of A without B $A \setminus B$

b is the number of connected components of B without A $(B \setminus A)$

o is the number of connected components of the outside area that does belong to neither A nor B $((A \cup B)^c)$

The relation "a" in Figure 2, for instance, corresponds to the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ as there is no overlap between the circular object A and the candy-cane shaped object B, and $A \setminus B, B \setminus A, (A \cup B)^c$ each consist of a single connected region. In contrast, the matrix for "h" is $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ as all relevant regions consist of a single component except for the complement of $(A \cup B)$ (everything that does belong to neither A nor B), which has two connected components. In total, this approach allows for distinguishing 23 "simple" modes of overlap relations for regions without holes (see Figure 2) in which the elements of the overlap matrix do not get higher than two. In principle, however, the approach can distinguish an infinite

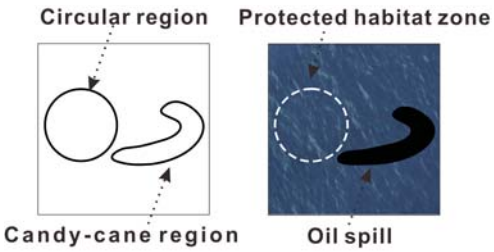

Fig. 1. Example of geometric (left) and semantic (right) stimuli used in the four experiments.

number of increasingly complex topological relations reflected by increasingly higher numbers in the matrix.

II. Methods

We used category construction tasks [36], also referred to as free classification [37], and employed our custom made software solution CatScan. Participants were presented with graphical stimuli that were created based on distinctions made in the modes of overlap calculus. Examples are provided in Figure 2 and details are discussed in Section II-B. Participants created categories based on their own assessment of the stimuli. In the first two experiments, the instructions did not contain any indication of how many groups should be created. In the other two experiments, participants were explicitly instructed to create as many meaningful groups as possible. The participants' category construction behavior of the last two experiments was then compared against 1) results from the first two experiments that did not require participants to create as many groups as possible and 2) categories (modes of overlap) that Galton's approach would differentiate. We were able to show in previous experiments [33] that semantics has an influence on how people conceptualize qualitative spatial relations such as those identified by the RCC and IM models. We therefore used both a semantic and a purely geometric version of the graphical stimuli each (see Figure 1 for an example) leading to the 2×2 experiments.

A. Participants

For each of the four experiments we recruited 20 participants through Amazon Mechanical Turk. Details are as follows[1]:

- **geometric figures + regular instructions:** 8 female, 12 male, average age 31.3
- **semantic figures + regular instructions:** 8 female, 12 male, average age 36.0
- **geometric figures + many group instructions:** 11 female, 9 male, average age 29.8
- **semantic figures + many group instructions:** 6 female, 14 male, average age 28.8

[1]One participant in the geometric + many groups experiment and two in the semantic + many groups experiment self-reported an age of 100 years and were excluded form the average age calculation.

B. Materials

Using the modes of overlap relations defined by Galton [1] (see Figure 2) as a template, we created two sets of icons showing either two geometric shapes or an oil spill together with a protected habitat zone (Figure 1). We made the following modifications to the original set of formally distinguished modes of overlap: First, we felt it to be important to include a disconnected relation in addition to the relation where the two regions are externally connected. Hence, the original relation "a" (see Figure 2) has been split into "a" for the disconnected case and "aa" for the externally connected case. These two situations are not distinguishable by Galton's overlap matrix but are distinguished in RCC-8 and Egenhofer's 9-intersection model. Second, we included a distinction between tangential and non-tangential proper part, again made in both RCC-8 and IM, that captures whether the contained object in relation "d" touches the boundary of the containing object or not. As a result, the original relation "d" has been split into "d" for non-tangential proper part and "dd" for tangential proper part. Third, we omitted the following original relations as they would have required significant changes in the visualization: relation "c", in which both regions have exactly the same spatial extension (which is difficult to visualize) and then relation "e", the inverse of "k", and relation "f", the inverse of "d", which would have required a complete reversal of the size differences between the two involved objects.

For each of the resulting 22 relations (Figure 2), we created four random variations in the following way: We produced a prototypical example configuration first, consisting of a circular region and a region in the form of a candy-cane, which allowed us to realize all relations without dramatic shape changes. The second, third, and forth variation were created by randomly rotating the prototypical example by an angle between 0-90 degrees, 90-180 degrees, and 180-270 degrees, respectively, followed by further scaling both regions down by a random factor between 0.8-1.0. Due to the constraints imposed by some overlap relations, the sizes of the candy-cane region are restricted to be relatively large in some cases (e.g., "i", "j", and "l") and relatively small in some others (e.g., "d", "dd", and "k"). For those overlap relations in which the candy-cane region can be scaled down by 50% without changing the relation, we did so for the first and second variation. In the instructions, however, we explicitly asked participants to ignore size. This was the only way to ensure that certain modes of overlap relations are not singled out simply because they are larger or smaller as size is potentially a strong grouping criterion [38]. Finally, all icons were checked to ensure that the spatial relation in each variation is perceptual unambiguous as we are focusing on human concepts of space.

C. Procedure

We created a stand-alone Java version of our category construction assessment software, CatScan (see Figure 3), such that it is compatible with AMT. Recruitment and payment was organized through AMT but participants downloaded CatScan and ran it locally on their computers. A results file is created at the end of each experiments which participants had to upload back to AMT. We used the available AMT command line tools to be able to automatically generate HIT descriptions with a running participant number. Participants read the general HIT

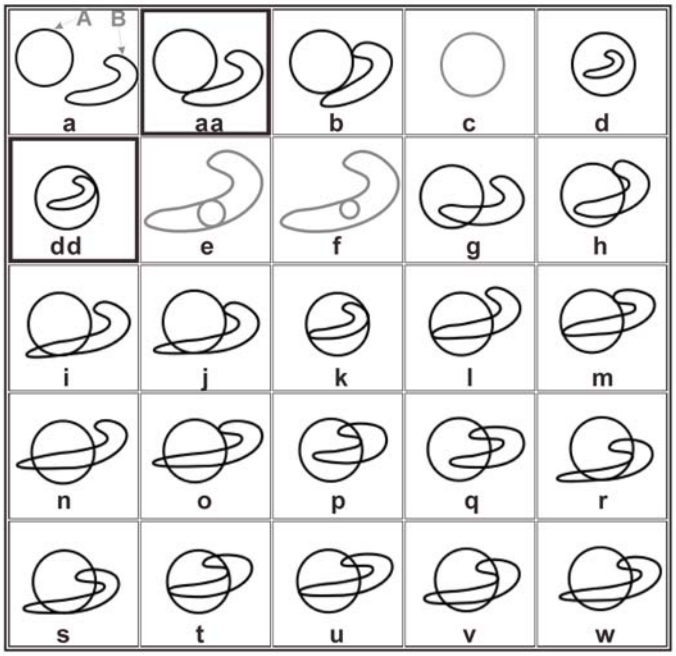

Fig. 2. Modes of overlap including all of Galton's 23 basic relations plus two additional relations added to these experiments ("aa" and "dd"). Relations in light gray were omitted from the experiments.

Fig. 3. CatScan Interface. Top: Initial screen; Bottom: Ongoing experiment.

instructions and had to enter the participant number into the CatScan interface. We performed several pre-tests to tailor our experiment and instructions to the specifics of AMT and Turkers' perspectives: a) We only allowed Turkers with a HIT approval rating of at least 95% to participate in this experiment; b) we split the payment into two components: Turkers received \$1.50 for their participation in the experiments and had the option to earn an additional \$1.50 as a bonus. They were informed that they would only receive the bonus in case they performed the category construction task thoroughly, label each category thoughtfully, and not use size as a grouping criterion. Each participant performed two tasks: a category construction task and a linguistic labeling task. All 88 overlap icons (4 for each of the 22 modes of overlap) for one of the two conditions (geometric or semantic) were initially displayed on the left side of the screen (Figure 3, top). Participants were required to create their own categories on the right side of the screen. After creating at least one empty category, they were able to drag and drop icons from the left side into a category on the right side (Figure 3, bottom). In the first two experiments, they were explicitly advised that there was no right or wrong answer regarding either the number of categories or which icons belong to which category (with the exception that they had to ignore size differences). In the other two conducted experiments, participants were additionally instructed to create as many meaningful groups as possible. Participants in all experiments had the opportunity to move icons between categories, move icons back to the left side, or delete whole categories. The main category construction experiment was preceded by a warm up task (sorting animals) to acquaint participants with the software. Participants performed a linguistic labeling task upon finishing the main category construction experiment. They were presented with the categories they had created and had to provide two linguistic descriptions: a short name of no more than five words and a longer description detailing their rationale(s) for placing icons into a particular category.

III. Results

In the analysis of the behavioral data collected, we will focus primarily on the category construction (conceptualization) behavior of participants in the four experiments: geometric figures + regular instructions (Geo), semantic figures + regular instructions (Sem), geometric figures + instructions to create as many meaningful groups as possible (GeoMany), and semantic figures + instructions to create as many meaningful groups as possible (SemMany). As a first analysis, comparing the number of groups participants created using one-way analysis of variance (ANOVA) showed that there are no statistically significant differences between the four experiments at the .05 confidence level. Likewise, comparing the time participants spent on the grouping task using ANOVA did not show any statistically significant differences either.

Figure 4 summarizes several aspects of multiple analyses we then performed on the participants' grouping behavior. It shows a combination of heat maps and dendrograms (resulting from applying Ward's clustering method) for the four experiments. The heat maps visualize the overall grouping behavior of all participants determined in the following way: A participant's grouping behavior was recorded in a 88×88 (88 is the number of icons in each experiment) similarity matrix in

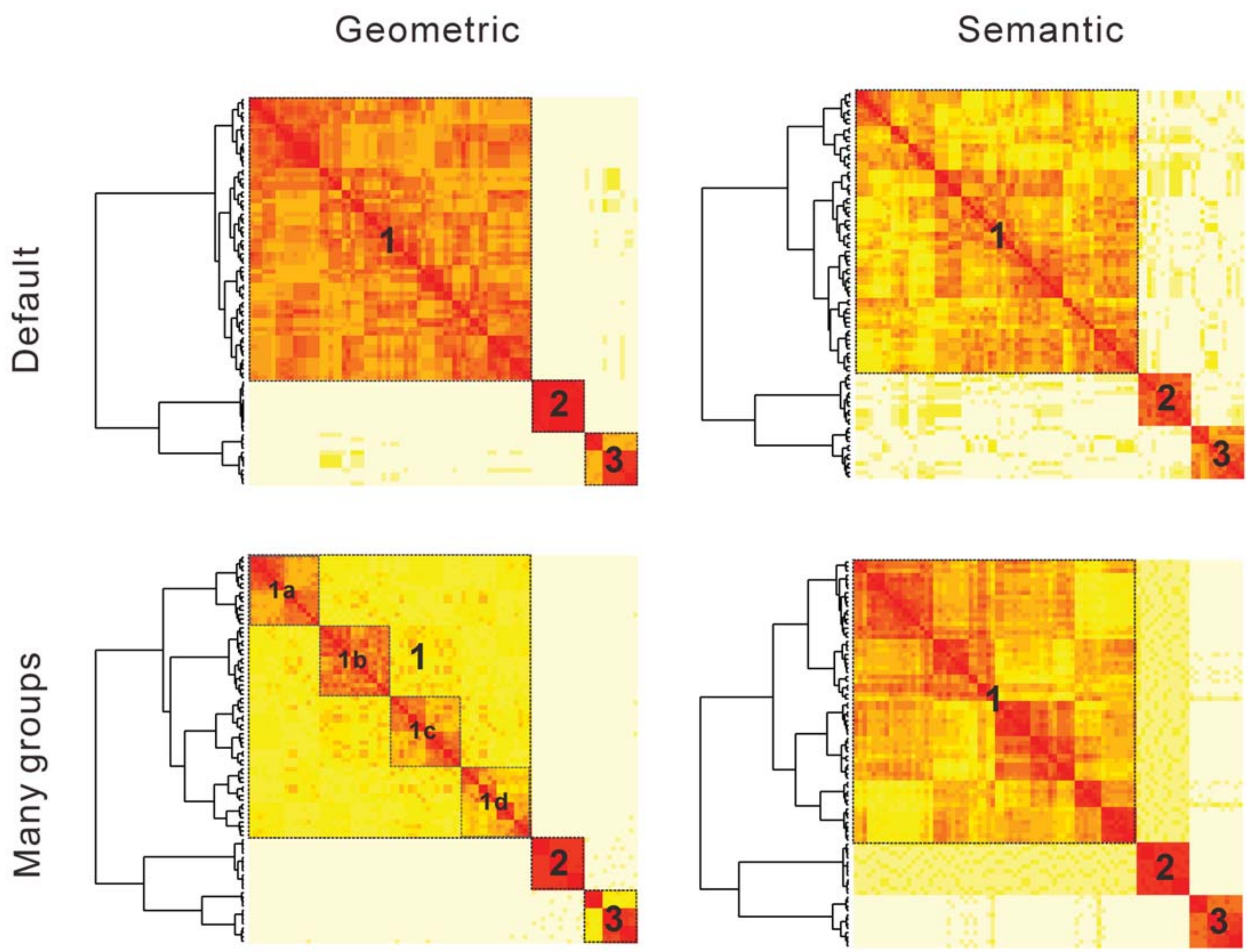

Fig. 4. Heat maps visualizing overall similarity assessment between pairs of icons for the four experiments with dendrograms resulting from Ward's method.

which a '0' indicates that two icons were not placed into the same group, while a 1' indicates that two icons were placed into the same group. An overall similarity matrix (OSM) is computed by summing up over the individual matrices of all 20 participants. Hence, similarity of two icons in the OSM is represented by a number ranging from 0 to 20. The heat maps in Figure 4 directly visualize the values in the OSM matrices using colors, white cells for 0 (least similar) and red for 20 (most similar) for the four experiments. The order of icons in rows and columns are adjusted based on the results of the cluster analysis using Ward's method such that very similar icons are alway neighbored to each other.

On a coarse level, a three-cluster structure can be found in all four experiments (see Figure 4). Cluster 1 contains all "g", "h", "i", "j", "l", "m", "n", "o", "p", "q", "r", "s", "t", "u", "v", and "w" icons. Cluster 2 contains all "d", "dd", and "k" icons. Cluster 3 contains all "a", "aa", and "b" icons. While Figure 4 is showing only results from Ward's method, we validated the results using different clustering methods (Ward's method, average linkage, and complete linkage), following recommendations by [39] and [40]. The results are summarized in Table I and can be described as follows:

TABLE I. COMPARISON OF CLUSTERING RESULTS USING DIFFERENT METHODS (WARD'S METHODS, AVERAGE LINKAGE, AND COMPLETE LINKAGE) . YES/NO INDICATE WHETHER THE CLUSTERS FOR A GIVEN NUMBER OF CLUSTERS ARE IDENTICAL OVER ALL METHODS.

Experiment	2 Clusters	3 Clusters	4 Clusters
Geo	No	Yes	No
Sem	No	Yes	No
GeoMany	Yes	Yes	Yes
SemMany	No	Yes	No

(1) We found strong indications that the mentioned three cluster solution is the most stable conceptual structure identified by participants across all four experiments independent of the semantics and the explicit instruction to create as many meaningful groups as appropriate. The three general relations distinguished in all four experiments are: non-overlapping configurations (Cluster 3), proper part relations (Cluster 2), and various modes of overlap (Cluster 1). Across all four experiments and all three clustering methods that we used

for validation purposes, the three cluster solution shows no "classification errors", that is, all three clusters each time contained the same icons.

(2) Across all four experiments the proper part relations, "d", "dd", and "k" are conceptually very close.

(3) Contextual factors such as changing the semantics of the relations involved and/or changing the instructions (i.e., explicitly asking participants to create as many meaningful groups as possible) does change the conceptual structure somewhat but does not change the big picture that the most intuitive distinction is indeed the one into non-overlapping, overlapping, and proper part relations. The most notable differences we found can be summarized as follows: The number of disconnected components of the intersection between the two regions becomes more relevant in the overall grouping behavior of the participants in the SemMany experiment. When the instruction to use as many meaningful groups as possible was present, those relations in which the intersection consists of a single connected region (i.e., "g", "h", "i", "j", "l", "m", "n", and "o") were distinguished from those relations in which the intersection consists of two disconnected regions (i.e., "p", "q", "r", "s", "t", "u", "v", and "w") in the resulting dendrograms for average linkage and Wards method (but not complete linkage). This difference is reflected in Galton's model by different values for the element x in the overlap matrix (see Section I-C) and, hence, can be seen as evidence that counting the number of disjoint components as in the Galton approach can play a role under specific conditions. In the geometric experiment we find a similar but more pronounced effect insofar as the undifferentiated overlap cluster in Geo is separated into four sub-clusters in GeoMany (see again Figure 4) and that this finer level of granularity is holding against cluster validation analysis, that is, it is consistent across the different clustering methods. The four sub-clusters are:

- **Sub-cluster 1a**: all "p", "q", "u" ,"t" icons
- **Sub-cluster 1b**: all "r", "s", "v", "w" icons
- **Sub-cluster 1c**: all "i", "j", "n", "o" icons
- **Sub-cluster 1d**: all "g", "h", "m", "l" icons

Interestingly, this distinction into four sub-clusters is reflected in Galton's approach by a combination of the values 1 and 2 for the x and b values in the overlap matrix (Section I-C). This means when encouraged to produce many groups, several participants adopted a similar approach as in Galton's modes of overlap calculus but only with regard to two of the four parameters in the overlap matrix. Notably this means that the number of components of $B \setminus A$ played a role, while that of $A \setminus B$ did not (with A being the circle and B being the candy-cane shaped object).

IV. Discussion

The results provide a number of important insights into questions of granularities of intuitive and commonsense conceptualizations of two spatially extended entities overlapping one another. The most important finding is that overall the distinctions made by the original topological calculi RCC and IM are well reflected in the results presented here, that is, the concept of overlap is a cognitively robust concept even if finer distinctions are possible. However, RCC and IM themselves are specified on two different levels of granularity with five and eight relations, respectively [34]. In the context of this article, it is worthwhile to look at these different levels for two reasons: First, we might find that one level is more intuitive than the other. Second, while the fine (eight relations) version of both RCC and IM distinguish the same topological relations, the coarse (five relations) versions do not. The coarse versions of RCC and IM agree on merging proper part relations (here: "d", "dd", and "k"). This agreement is nicely reflected in the behavioral data as Figure 4 shows clearly that the proper part cluster is highly prominent across all four experiments (and all clustering methods, not visualized). This finding is also in line with research we conducted on topologically characterized movement patterns [12] and complex spatial relations [41]. RCC and IM disagree on whether disconnected and externally connected relations (here: "a", "aa", and "b") or externally connected and partial overlap relations should be merged. It is clear from our results that overlap relations are separated from externally connected relations, however, in the geometric experiments and especially in combination with the many group instructions the conceptual glue between "a" relations, on the one hand, and "aa" and "b" relations, on the other, is weakened.

One critical finding sheds light on the role of contextual factors such as adding a semantic context to the characterization of spatial configurations. Up to now, there is no clear theory that would allow for predicting the influence of semantics on how spatial configurations are conceptualized. There are a number of approaches such as Coventry and Garrod's *extra-geometric functional framework* [42], Talmy's work on *force dynamics* [43], or Regier's work on the *human semantic potential* [44], but they do not allow for making predictions easily on whether certain spatial relations identified in spatial calculi are conceptually similar or different [45]. Our own work has yielded results which underline the need for a more sophisticated theory of the semantics of spatial concepts. For example, in a recent set of animated experiments [46] we were able to show that the externally-connected relation (here: "aa", "b") is considered conceptually very different from the disconnected relation (here: "a") in case the involved objects are an island and an oil spill. In contrast, the results obtained in the present experiments do not show this differentiation in either semantic experiment. This can be explained by the different characteristics of the reference entities (island versus protected habitat zone) and the semantic association that externally connected, in case of an oil spill making it to the shore, is the start of a disaster but in a static snapshot and with no processes directly related to the border at a habitat zone this distinction is not prominent for the habitat + oil spill pair.

Most interesting we find though that in the "many groups" experiments the semantic context glues together spatial relations especially when compared to purely geometric configurations. This is impressively demonstrated by the consistent finer granularities that surface in the geometric version of the "many groups" experiments. In other words, on purely geometric grounds participants make finer distinctions of modes of overlap when asked to do so, while this is not the case in the experiments with a semantic context.

V. Conclusions

Granularities, changes in granularity, and intuitive levels of granularities of spatial relations are essential for manifold cognitive computing efforts. The fundamental concept of space we addressed in this article is the overlap relation between two spatially extended entities. Overlap relations are identified in the most prominent qualitative spatial calculi (e.g., RCC and IM), but are treated as a single homogeneous concept. Galton's modes of overlap calculus [1] accounts for potentially important distinctions within the concept of overlap, which could be relevant for both spatial thinking and for geospatial applications. While RCC does not easily allow for a finer distinction of different modes of overlap, the intersection model of Egenhofer's approach lends itself for rendering qualitative distinctions between different modes of overlap more precise by designing an overlap matrix notation. Each mode of overlap is bounded by salient discontinuities in the formal qualitative description, that is, it can be considered invariant until changes in the configuration surface in changing values in the overlap matrix. In other words, it is assumed that the characterization captured by a specific combination of values in the overlap matrix identifies an infinite number of situations by singling out the essential characteristics under which they are identical (such as the four variations we created for every mode of overlap).

While some work in the area of event segmentation addresses issues of finer or coarser segmentation strategies [47], there have not been many approaches to explore the different levels of granularity for spatial relations from a cognitive perspective. We therefore combined four experiments in this article that we ran as a combination of our custom-made category construction software, CatScan, and the crowdsourcing platform Amazon Mechanical Turk. Two experiments in which we asked participants to create groups according to whatever criteria they deemed appropriate and two experiments using the same stimuli (geometric figures versus a semantic scenario with an oil spill and a protected habitat) but with the changed instruction to create as *many* meaningful categories as possible. The take home message from comparing results of all four experiments is that the single overlap concept is very robust. While there are finer levels of granularity present, partially following Galton's approach of counting components of the different regions but not giving equal weight to all four parameters in the overlap matrix, these do not surface as being salient enough across all participants and conditions.

A couple of concluding remarks on our approach and the bigger picture: We are aiming for designing an efficient research framework in which it becomes possible to calibrate spatio-temporal calculi against behavioral data. These calculi are often used at the interface of humans and machines such as geo-spatial search engines [48], [49]. Efficiency of collecting behavioral data is increased through crowdsourcing platforms such as Amazon's Mechanical Turk, which ideally supplement laboratory experiments and provide a number of advantages: (a) participants cover a wider range of the population, that is, they are not only students; (b) data can often be collected over night; and (c) while we decided to pay on the higher end of what is expected within AMT, we were able to run the same experiments at about a third of the costs. Finally, after some tests, we are, most importantly, confident that we have tailored our experimental setup to the specifics of AMT such that results are comparable to laboratory experiments. We are, however, not only gaining efficiency by using a crowdsourcing platform but also by having a software that collects all experimental data efficiently and, importantly, analyses tools that allow for instantly analyzing at least the most basic data. We consider this an essential step toward cognitively grounding formal approaches designed for the interface of humans and machines but essentially for all kinds of classification and categorization tasks especially those that compare algorithmic and formal solutions to human behavior and cognition [50].

Acknowledgment

This research is funded by the National Science Foundation under grant #0924534.

References

[1] A. Galton, "Modes of overlap," *Journal of Visual Languages and Computing*, vol. 9, pp. 61–79, 1998.

[2] J. R. Hobbs, "Granularity," in *Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Angeles, CA, August 1985*, A. K. Joshi, Ed. San Francisco and CA: Morgan Kaufmann, 1985, pp. 432–435.

[3] B. Kuipers, "The spatial semantic hierarchy," *Artificial Intelligence*, vol. 119, no. 1-2, pp. 191–233, 2000.

[4] K.-F. Richter and A. Klippel, "A model for context-specific route directions." in *Spatial Cognition IV. Reasoning, Action, Interaction*, C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, Eds. Berlin: Springer, 2005, pp. 58–78.

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

[6] M. J. Egenhofer and R. D. Franzosa, "Point-set topological spatial relations," *International Journal of Geographical Information Systems*, vol. 5, no. 2, pp. 161–174, 1991.

[7] J. Renz and D. Mitra, "Qualitative direction calculi with arbitrary granularity," in *PRICAI 2004: Trends in Artificial Intelligence, 8th Pacific Rim International Conference on Artificial Intelligence, Proceedings*. Springer, 2004, pp. 65–74.

[8] R. Moratz and J. O. Wallgrün, "Spatial reasoning with augmented points: Extending cardinal directions with local distances," *Journal of Spatial Information Science (JOSIS)*, no. 5, pp. 1–30, 2012.

[9] A. Clark, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing.* Cambridge and MA: MIT Press, 1989.

[10] W. Kuhn, "An image-schematic account of spatial categories," in *Spatial Information Theory*, S. Winter, B. Kuipers, M. Duckham, and L. Kulik, Eds. Berlin: Springer, 2007, pp. 152–168.

[11] G. Lakoff and M. Johnson, *Metaphors we Live by*. Chicago: University of Chicago Press, 1980.

[12] A. Klippel, R. Li, J. Yang, F. Hardisty, and S. Xu, "The Egenhofer-Cohn hypothesis: Or, topological relativity?" in *Cognitive and Linguistic Aspects of Geographic Space - New Perspectives on Geographic Information Research*, M. Raubal, A. U. Frank, and D. M. Mark, Eds. Berlin: Springer, 2013, pp. 195–215.

[13] D. M. Mark, "Spatial representation: A cognitive view," in *Geographical information systems: Principles and applications*, P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, Eds. New York: Wiley, 1999, pp. 81–89.

[14] M. Worboys and M. Duckham, *GIS: A Computing Perspective.*, 2nd ed. Boca Raton FL: CRC Press, 2004.

[15] D. J. Peuquet, *Representations of space and time*. New York NY: Guilford Press, 2002.

[16] F. Klein, "Vergleichende Betrachtungen über neuere geometrische Forschungen ("A comparative review of recent researches in geometry")," *Mathematische Annalen*, vol. 43, pp. 63–100, 1872.

[17] J. M. Mandler, "How to build a baby II. Conceptual primitives," *Psychological Review*, vol. 99, no. 4, pp. 587–604, 1992.

[18] F. Klix, *Die Natur des Verstandes*. Göttingen: Hogrefe, 1992.

[19] R. Shaw, M. McIntyre, and W. Mace, "The role of symmetry in event perception," in *Perception*, R. B. MacLeod and H. L. Pick, Eds. Ithaca: Cornell University Press, 1974, pp. 276–310.

[20] J. Gibson, *The ecological approach to visual perception*. Boston and MA: Houghton Mifflin, 1979.

[21] G. Lakoff, "The invariance hypothesis: is abstract reason based on image schemata?" *Cognitive Linguistics*, vol. 1, no. 1, pp. 39–74, 1990.

[22] M. J. Maguire, J. Brumberg, M. Ennis, and T. F. Shipley, "Similarities in object and event segmentation: A geometric approach to event path segmentation," *Spatial Cognition and Computation*, no. 3, pp. 254–279, 2011.

[23] A. Galton, *Qualitative spatial change*, ser. Spatial information systems. Oxford: Oxford Univ. Press, 2000.

[24] S. Scheider, K. Janowicz, and W. Kuhn, "Grounding geographic categories in the meaningful environment," in *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings*, K. Stewart Hornsby, C. Claramunt, M. Denis, and G. Ligozat, Eds. Berlin: Springer, 2009.

[25] M. J. Egenhofer, "The family of conceptual neighborhood graphs for region-region relations," in *Sixth International Conference, GIScience 2010 Zürich, Switzerland, Sep. 14-17 2010, Proceedings*, S. I. Fabrikant, T. Reichenbacher, M. van Krefeld, and C. Schlieder, Eds. Berlin: Springer, 2010, pp. 42–55.

[26] Y. Kurata and M. J. Egenhofer, "Interpretation of behaviors from a viewpoint of topology," in *Behaviour monitoring and interpretation*, B. Gottfried and H. Aghajan, Eds. Amsterdam: IOS Press, 2009, pp. 75–97.

[27] B. Smith and D. M. Mark, "Geographical categories: An ontological investigation," *International Journal of Geographical Information Science*, vol. 15, no. 7, pp. 591–612, 2001.

[28] A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing user studies with Mechanical Turk," in *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. New York and NY and USA: ACM, 2008, pp. 453–456.

[29] J. Heer and M. Bostock, "Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design," in *ACM CHI 2010*, 2010.

[30] D. G. Rand, "The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments: Evolution of Cooperation," *Journal of Theoretical Biology*, vol. 299, no. 0, pp. 172–179, 2012.

[31] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgement and Decision Making*, vol. 5, no. 5, pp. 411–419, 2010.

[32] A. G. Cohn and J. Renz, "Qualitative spatial representation and reasoning," in *Handbook of knowledge representation*, ser. Foundations of artificial intelligence, F. van Harmelen, V. Lifschitz, and B. Porter, Eds. Amsterdam: Elsevier, 2008, pp. 551–596.

[33] A. Klippel, "Spatial information theory meets spatial thinking - Is topology the Rosetta Stone of spatio-temporal cognition?" *Annals of the Association of American Geographers*, vol. 102, no. 6, pp. 1310–1328, 2012.

[34] M. Knauff, R. Rauh, and J. Renz, "A cognitive assessment of topological spatial relations: Results from an empirical investigation," in *Spatial information theory: A theoretical basis for GIS*, S. C. Hirtle and A. U. Frank, Eds. Berlin: Springer, 1997, pp. 193–206.

[35] D. M. Mark and M. J. Egenhofer, "Topology of prototypical spatial relations between lines and regions in English and Spanish," in *Proceedings, Auto Carto 12, Charlotte, North Carolina, March 1995,*, 1995, pp. 245–254.

[36] D. L. Medin, W. D. Wattenmaker, and S. E. Hampson, "Family resemblance, conceptual cohesiveness, and category construction," *Cognitive Psychology*, vol. 19, no. 2, pp. 242–279, 1987.

[37] E. M. Pothos and N. Chater, "A simplicity principle in unsupervised human categorization," *Cognitive Science*, vol. 26, no. 3, pp. 303–343, 2002.

[38] A. Klippel, R. Li, F. Hardisty, and C. Weaver, "Cognitive invariants of geographic event conceptualization: What matters and what refines," in *Sixth International Conference, GIScience 2010 Zürich, Switzerland, Sep. 14-17 2010, Proceedings*, S. I. Fabrikant, T. Reichenbacher, M. van Krefeld, and C. Schlieder, Eds. Berlin: Springer, 2010, pp. 130–144.

[39] A. J. Kos and C. Psenicka, "Measuring cluster similarity across methods," *Psychological Reports*, vol. 86, pp. 858–862, 2000.

[40] D. J. Ketchen and T. M. Hult, "Validating cluster assignments," *Psychological Reports*, vol. 87, pp. 1057–1058, 2000.

[41] J. O. Wallgrün, J. Yang, A. Klippel, and F. Dylla, "Investigations into the cognitive conceptualization and similarity assessment of spatial scenes," in *Geographic Information Science*, N. Xiao, M.-P. Kwan, M. F. Goodchild, and S. Shekhar, Eds. Berlin: Springer, 2012, pp. 212–225.

[42] K. R. Coventry and S. Garrod, *Saying, Seeing, and Acting: The Psychological Semantics of Spatial Prepositions*. Hove: Psychology Press, 2004.

[43] L. Talmy, "Force dynamics in language and cognition," *Cognitive Science*, vol. 12, no. 1, pp. 49–100, 1988.

[44] T. Regier, *The human semantic potential: Spatial language and constraint connectionism*. Cambridge (MA) and London: The MIT Press, 1996.

[45] B. Li and F. Fonseca, "TDD: A comprehensive model for qualitative spatial similarity assessment," *Spatial Cognition and Computation*, vol. 6, no. 1, pp. 31–62, 2006.

[46] J. Yang, A. Klippel, and R. Li, "Cognitive saliency of topological change under expansion and contraction," *International Journal of Geographical Information Science*, in revision.

[47] J. M. Zacks, B. Tversky, and G. Iyer, "Perceiving, remembering, and communicating structure in events." *Journal of Experimental Psychology: General*, vol. 130, no. 1, pp. 29–58, 2001.

[48] M. J. Egenhofer, "Spatial SQl: A query and presentation language," *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, pp. 86–95, 1994.

[49] A. I. Abdelmoty and B. A. El-Geresy, "Modeling spatial relations between places to support geographic information retrieval," in *Knowledge-Based Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, I. Lovrek, R. J. Howlett, and L. C. Jain, Eds. Springer Berlin Heidelberg, 2008, vol. 5178, pp. 689–694.

[50] Z. Ul-Qayyum, A. G. Cohn, and A. Klippel, "Psychophysical evaluation for a qualitative semantic image categorisation and retrieval approach," in *Trends in applied intelligent systems*, ser. Lecture notes in computer science Lecture notes in artificial intelligence, N. García-Pedrajas, F. Herrera, C. Fyfe, J. M. Benítez, and M. Ali, Eds. Berlin: Springer, 2010, vol. 6098, pp. 321–331.