to appear in proceedings of COSIT 2013

Fundamental Cognitive Concepts of Space (and Time): Using Cross-Linguistic, Crowdsourced Data to Cognitively Calibrate Modes of Overlap

Alexander Klippel[1]*, Jan Oliver Wallgrün[1], Jinlong Yang[1], Jennifer S. Mason[1], Eun-Kyeong Kim[1], and David M. Mark[2]

[1] Department of Geography, GeoVISTA Center, The Pennsylvania State University
University Park, PA 16802, USA
{klippel,wallgrun,jinlong,jms1186,eun-kyeong.kim}@psu.edu
http://www.cognitiveGIScience.psu.edu

[2] Department of Geography, State University of New York at Buffalo
Buffalo, NY 14261-0023, USA
dmark@buffalo.edu

Abstract. This article makes several contributions to research on fundamental spatial and temporal concepts: First, we set out to render the notion of fundamental concepts of space and time more precise. Second, we introduce an efficient approach for collecting behavioral data combining crowdsourcing technology, efficient experimental software tools, and an effective and comprehensive analysis methodology. Third, we present behavioral studies that allow for identifying and calibrating potential candidates of fundamental spatial concepts from a cognitive perspective. Fourth, one prominent topic in the area of spatio-temporal cognition is the influence of language on how humans conceptualize their dynamic spatial environments. We used the aforementioned framework to collect data not only from English speaking participants but also from native Chinese and Korean speakers. Our application domain are the *modes of overlap* proposed by Galton [13]. We are able to show that the originally proposed spatial relations of the region connection calculus and intersection models are capturing cognitively fundamental distinctions that humans make with respect to modes of overlap. While finer distinctions are formally possible, they should not be considered fundamental conceptualizations in either Chinese, Korean, or English. The results show that our framework allows for efficiently answering questions about fundamental concepts of space, time, and space-time essential for theories of spatial information.

Keywords: Fundamental concepts of space and time, qualitative spatio-temporal reasoning, linguistic relativity, crowdsourcing

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

"To study concepts, cognitive [spatial] scientists must first identify some." (Malt et al. 2011 [45])

1 Introduction

The Conference on Spatial Information Theory (COSIT) has a long tradition of addressing deeply theoretical questions of space and time, often from a cognitive perspective. In this spirit, we concentrate on the notion of concepts fundamental to the human understanding of space and time. We use the term fundamental loosely indicating concepts at the core of how humans are making sense of spatio-temporal information; alternative terms may be *intuitive*, *naive* [12, 19], *common sense*, or *spatial conceptual primitives* [47]. We will use the acronym Func2st to indicate fundamental cognitive concepts of space, time, and space-time. Critical questions for cognitively oriented spatial scientists are how Func2st can be identified and defined, whether formally hypothesized Func2st have cognitive counterparts (i.e., are cognitively adequate [33]), how Func2st are related to language, and, importantly, whether Func2st in one language correspond those in another language. This research relates to work on conceptual primitives [70, 2], theories of image schemata [24, 37, 46], the spatial foundations of the human conceptual system [47], and the relation between concepts and words [22, 71]. The importance of such research is manifold as it adds to the growing body of knowledge that addresses questions of linguistic relativity [17, 71], it adds to the pursuit of defining the foundations of meaning (from a spatio-temporal perspective) [64], and results of this research potentially improve approaches for transforming data into knowledge, for instance when interpreting data from sensor networks as in [72]. The latter aspect is also important for cognitive scientists. An example is the continuous to discrete transformation essential for language production [16]. Continuously varying information needs to be translated into quasi-symbolic equivalence classes that allow for producing reliable behavior.

To answer the questions raised above, this article is structured as follows. First, we provide a background discussion on relevant topics: fundamental spatial and spatio-temporal concepts, crowdsourcing, and cross-linguistic differences. Second, we provide details on an experiment that uses Galton's [13] *modes of overlap* to demonstrate the overall approach we propose (comparing category construction behavior of English, Chinese, and Korean speakers) to shed light on questions surrounding Func2st. Third, we discuss results, present conclusions, and describe future research opportunities.

2 Background

2.1 Func2st and their behavioral evaluation

The term *fundamental spatial concepts* is not well defined [73, 56, 37] (see also Table 1). However, we can use the notion of *invariance* to render it more precise—across disciplines. Researchers in many scientific fields from the cognitive to the spatial sciences have addressed the topic of invariance in the context of cognitive information processing as well as to formally distinguish fundamental concepts of of space and time. Worboys and Duckham [73], for example, utilize the concept of invariance as a framework for their chapter on Fundamental Spatial Concepts. Proposed first by Felix Klein (see Erlangen program [26]), geometries can be distinguished based on invariant properties under certain transformations. This approach allows for differentiating Euclidean geometry from set-based geometry as well as topology which can be seen as "rubber-sheet geometry" concerned with properties that remain invariant under topological transformations such as bending and stretching. Paralleling formal approaches, perceptual and cognitive invariants, which we also find to be associated with conceptual primitives (e.g., [46]), have long been of interest to the cognitive science community. Klix [32] stated that the human mind, in adapting to its environment, identifies invariant characteristics of information that form the basis for cognitive processes. Shaw [65] used 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. Last, the classic work by Gibson [15] refers to temporally constant characteristics of environments as structural invariants. The basic work of these scholars is featured in many recent approaches utilizing the concept of invariance to explain, for example, perception, categorization, and event cognition [38, 44, 18, 64, 16]. The importance of identifying invariants of environmental information is prominently noted by Galton [14], who speaks of our ability to intersubjectively identify invariants of space and time which allow us to construct a shared understanding of our physical (and social) environments. Without the agreement that certain characteristics of spatial environments ground our meaningful understanding of spatial environments (e.g., [63]), the concept of a shared reality and our ability to communicate about this reality would not be possible. Many researchers have pointed to topology as the cognitively most important qualitative formal theory that allows for rendering the notion of invariance more precise (see [68, 10, 35, 36, 66, 32]). While topology is unquestionably important for understanding cognition, there are other theories that identify invariants potentially relevant for humans' fundamental understanding of space and time.

Table 1 provides an overview of a first collection of what could be considered fundamental spatio-temporal concepts extracted from the relevant literature. Please note that this list does not claim to be comprehensive at this point. This list can and should be extended to capture concepts deemed fundamental for thinking spatially and temporally by researchers from both the cognitive and spatial sciences (see Section 5). The table also lists some cognitive evaluations that have been conducted to prove, disprove, or modify the cognitive adequacy of formal models attempting to better capture human understanding and processing of formally identified fundamental concepts.

Over the last years, members of the Human Factors in GIScience Lab[3] have established an efficient framework (see Figure 1) to conduct behavioral experiments with the goal to evaluate Func2st using category construction tasks [52]. In contrast to many linguistic-driven approaches (see Section 5.2 for a discus-

[3] www.cognitiveGIScience.psu.edu

Table 1. Candidates for fundamental spatial, temporal, and spatio-temporal concepts

Spatio-temporal concepts	Short characterization	Cognitive calibration
Topology [60, 11]	Different calculi that address the most basic properties of space, such as connectedness.	Static: [33, 48] Dynamic: [27]
Modes of overlap [13]	Formal extension of the single overlap relation identified in topological calculi to capture more complex forms of overlap (see Figure 2).	Pre-study [69]
Perceptual topology (e.g., [8])	Real world relations may be perceived as being topological but formally they may not be: Example: a road enters a housing estate albeit the estate is constituted by individual houses.	No assessment
Allen's interval algebra (1D topology) [1]	Definition of possible relations between time intervals (events), six relations plus inverse relations and the identity of two intervals.	[43]
Renolen's basic types of change [61]	Seven basic types of change: creation, alteration, destruction, reincarnation, merging/annexation, splitting/deduction, reallocation.	No assessment
Direction calculi (e.g., [42])	Modeling of directions as sectors or projections, differentiating cardinal and relative directions.	[29]
Chorematic Modeling [5]	Theory established by a French geographer Roger Brunet detailing 26 basic models to characterize spatial relations and processes.	No assessment
Shape (e.g., [14])	Shape of objects including trajectories.	(e.g., [54])

sion) to understanding human concepts, this framework uses qualitative formal theories as a hypothesis for salient discontinuities of space and time (see also [48, 33]). Given a formal model which categorizes relations over a fundamental spatial or temporal aspect into a finite (and typically small) number of equivalence classes, hereafter referred to as qualitative equivalence classes (QEC), a set of icons is generated consisting of an equal number of visual stimuli for each QEC. Human participants then group the icons (construct categories) in our own experimental software, CatScan, and the results can be analyzed using a comprehensive analysis methodology. Results from comparing the QEC-implied categories with the cognitive behavior shown in the experiment allows for drawing conclusions about the adequacy of the investigated formal model and may also lead to refined or improved (calibrated) models that, in turn, can improve spatio-temporal data processing applications. Additionally, insights are gained of how space and time are cognitively processed. More details on the involved tools and methodology will be given in later sections of this paper.

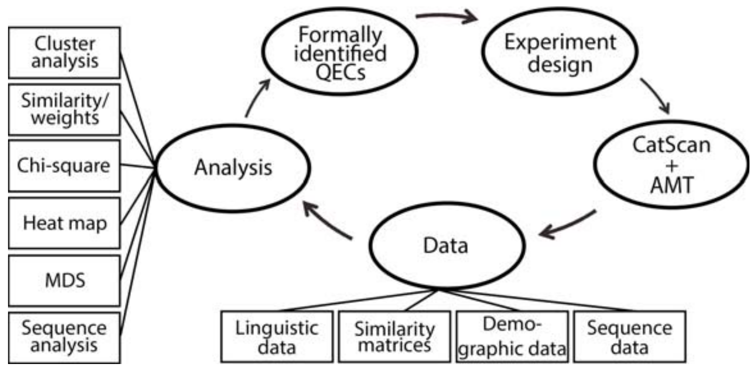

Fig. 1. Research framework (modified after [30]).

2.2 Crowdsourcing

Crowdsourcing has the potential to change and substantially advance the way that knowledge about Func2st is elicited. Multiple crowdsourcing platforms (e.g., Survey Monkey, Lime Survey) have been developed over the last decade and are now central to scientific and especially behavioral research (e.g., [25]). Of the many options that exist, Amazon's Mechanical Turk (AMT) has attracted a lot of attention and has developed into a serious alternative to using laboratory experiments (e.g., [20, 59, 50]). One of the main advantages is the more diverse (in comparison to only college students) pool of participants that can be accessed [62]. Most commonly, simple tasks are outsourced to a network of registered users of AMT. These tasks are referred to as HITs (Human Intelligence Tasks) and AMT guarantees that the tasks are performed anonymously. "Turkers" receive payment directly through AMT. To control the quality of responses, Turkers will only receive reimbursement when they perform a task satisfactorily. Additionally, restrictions can be placed by allowing, for example, only Turkers above a certain HIT approval rate, obtained through successfully completing experiments. Several studies indicate that results are comparable to regular laboratory experiments (e.g., [55, 59]). Some pre-tests that we conducted, however, suggest some cautions: In preparation for this article, we performed several (pre-)studies using AMT that allowed us to advise strategies to ensure the quality of crowdsourced data. It is essential to think of the experiments from the perspective of a Turker, that is, to maximize financial gain by finding the fastest yet correct way to perform a task. This leads to certain strategies that can be positive and/or negative. The bottom line is that Turkers' strategies have to be taken explicitly into account and channeled.

2.3 Cross-linguistic differences

While linguistic analysis as a window to cognition allows us to shed light on the question of underlying conceptual structures, the influence of language on the conceptualization of spatial relations (and on cognition in general) is a subject of ongoing debate [3, 9, 21, 39, 53][34]. There are various perspectives on the role of language in cognition that range from assuming universal conceptual categories that are essentially uniform across languages [57, 41] to assuming language dependent concepts [40] via recent approaches that propose a mapping between concepts and words in form of word-clusters [22]. A majority of researchers probably subscribe to the idea that language has some influence on cognition, specifically on the conceptualization of spatial relations, even though it is not completely deterministic. These theories are often discussed under the term *linguistic relativity* (e.g., [17, 71]). The question, however, is how profound this influence is and how it is manifested. One of the major research approaches to addressing this debate is to compare different languages in their expression of spatial relations and how their expressions may in turn influence nonlinguistic thought (see, for example, [4] and a reply by January and Kako [23]).

Instead of starting with identified linguistic differences, our approach allows for using formally identified spatial and temporal invariants as a hypothesis of Func2st across different languages. While research has demonstrated potentially different influences of the three languages (English, Chinese, and Korean) that we chose on the conceptualization of space and time (see [4, 51]), the level of detail with respect to formally identified invariants is often insufficient to derive exact hypotheses.

3 Methods

Revealing conceptual structures underlying humans understanding of space and time can be ideally addressed using category construction tasks [52], also referred to as conceptualization or free classification [58]. We have refined our custom-made software solution, CatScan, and extended it such that it is possible to run it in conjunction with Amazon Mechanical Turk. In the experiments, participants were presented with graphical stimuli that were created based on distinctions made by Galton's approach, exemplarily visualized in Figure 2 (details are given below). Participants created categories based on their own assessment of the stimuli. Participants' category construction behavior was subsequently compared against categories (modes of overlap) identified in Galton's approach. While we were able to show in previous experiments [27] that semantics has an influence on how people conceptualize spatial relations such as those identified by the region connection calculus (RCC) [60] and intersection models (IM) [11], we did not find any significant distinctions for the modes of overlap comparing a purely geometric scenario with the one involving an oil spill and protected habitat we used in this paper [69]. No sophisticated theory currently exists that predicts the influence of semantics on the conceptualization of qualitative spatial relations; we will return to this aspect in Section 6. The focus of this article is on whether someone's native language alters the conceptualization of spatial relations. Hence, we selected the oil spill plus protected habitat zone scenario as it does yield the same results as a purely geometric version.

Preliminary: Modes of overlap. The formalism investigated in this paper is the *modes of overlap* model by Galton [13]. This approach has been developed in response to a somewhat indifferent or unspecific treatment of the concept of overlap in the most important qualitative spatial calculi dealing with topological relations: RCC and IM. In summary: on a coarse level, RCC and IM both distinguish the same eight spatial relations between two spatially extended entities. A subset of these relations is present in the table shown in Figure 2 (`a`, `aa`, `c`, `d`, `dd`, `g`; for extensive treatments see [7]). These topology-based distinctions of spatial relations have been extensively analyzed in behavioral studies (e.g., [27, 33, 49]). Galton's motivation for developing a more differentiated picture was that configurations such as `g` and `q` (see Figure 2) cannot readily be distinguished in RCC and IM. He therefore developed an overlap matrix that would capture different modes of overlap based on the number of connected components of regions with overlap properties: For two regions A and B, the 2 x 2 matrix

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

contains the following information:

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 belongs to neither A nor B $((A \cup B)^c)$

For instance, the relation `a` in Figure 2 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. Without going into any more detail (please see [13]), this approach allows for distinguishing 23 simple modes of overlap for regions without holes (see Figure 2). While the approach can distinguish an infinite number of increasingly complex topological relations, these 23 modes of overlap are the most basic cases where no number in the matrix gets higher than two. We use a slightly modified version (explained below) of this set of relations as a hypothesis to test whether humans naturally make similar distinctions regarding modes of overlap between two regions, a topic that has not been evaluated from a cognitive-behavioral perspective.

Participants. Sixty-eight participants were recruited through Amazon Mechanical Turk to participate in three experiments (i.e., English, Chinese, and Korean). Eight participants across all three language versions were excluded because they

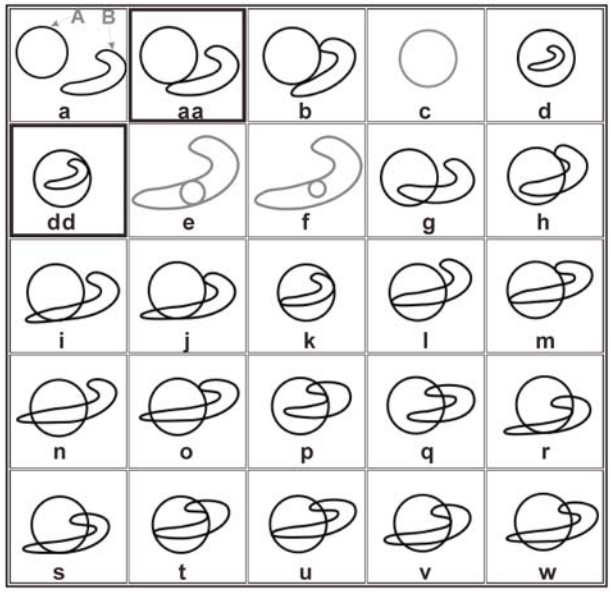

Fig. 2. Modes of overlap: 23 original modes of overlap relations between two spatially extended entities and modifications for the experimental setup used here. Two relations were added (aa, dd), three were omitted (c, e, f).

did not follow the instructions. Each language had 20 participants in the end. The English and Chinese version both included 8 females with an average age of 35.95 and 28.65, respectively. The Korean version had 10 female participants (average age: 30.25).

Materials. Based on the different modes of overlap defined by Galton [13] (see Figure 2), we created a set of icons showing different relations of an oil spill and a protected habitat zone. The icons show 22 different relations (see Figure 2) that resulted from making the following modifications to the original set of formally distinguished modes of overlap: First, we felt it is important to include a disconnected relation in addition to the relation where the two regions are only externally connected (see Figure 2, a and aa). Hence, the original relation a 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 and IM. Second, we wanted to include the distinction between tangential and non-tangential proper part also made in RCC and IM capturing whether or not the contained object in relation d touches the boundary of the containing object. 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 entities have exactly the same spatial extension, relation `e`, the inverse of `k`, which would have required a reversal of the size differences between the two involved regions, and relation `f`, the inverse of `d` which we omitted for the same reason as `e`.

The full icon set was created by generating four random variations of an initial prototype consisting of a circular region and a region in the form of a candy-cane for each of the 22 relations, which allowed us to realize all relations without dramatic shape changes. The second, third, and fourth variations were created by randomly rotating the prototypical example at an angle between 0-90 degrees, 90-180 degrees, and 180-270 degrees, respectively; next, both regions were further scaled down by a random factor between 0.8-1.0. Due to the constraints imposed by some overlap relations, the size of the candy-cane region was 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. Considering that size is potentially a strong grouping criterion [28], this was the only way to ensure that certain modes of overlap are not singled out simply because they are larger or smaller. Finally, all icons were checked to ensure that the spatial relation in each variation is perceptually unambiguous as we are focusing on human concepts of space.

Procedure. We modified CatScan, our category construction assessment software, to enable compatibility with AMT. For this purpose, we created a standalone Java version that can be downloaded as a .jar file from a server and run locally on participants' computers. Recruitment and payment was organized through AMT. We used the available command line tools to automatically generate HIT descriptions with a running participant number. Participants read the general HIT instructions and entered the participant number into the interface. At the end of the experiment, CatScan generated a result file with this unique ID, which participants uploaded to AMT. 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. As we encountered problems recruiting sufficient participants for the Korean experiment, we incrementally increased the payment to $2.50 + $1 bonus and lowered the required qualitatication to 50% with mixed results (see discussion in Section 5.3). Participants were informed that they would only receive the bonus if 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 (22×4) were initially displayed on the left panel of the screen (Figure 3, top). Participants were required to create categories on the right panel of the screen. After creating at least one empty category, they were able to drag icons from the left panel into a category on the right panel (Figure 3, bottom). They were explicitly advised that there was no right or wrong answer regarding the number of categories or which icons belong to which category (with the exception that they had to ignore size differences). They also had the opportunity to move icons between categories, move icons back to the left panel, or delete whole categories, in which case icons are placed back in the window on the left panel. The main category construction experiment was preceded by a warm up task (grouping animals) to acquaint participants with the software and category task. Participants performed a linguistic labeling task upon finishing the main category construction experiment. They were presented with the categories they created and provided 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.

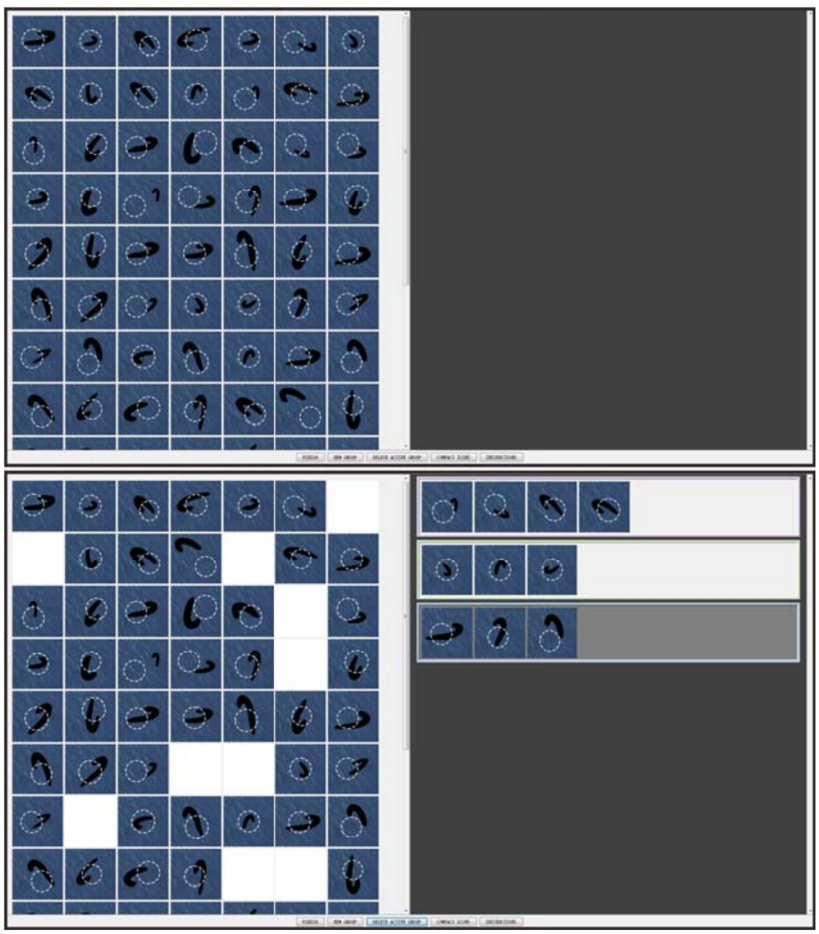

Fig. 3. CatScan interface. Top: initial screen; bottom: ongoing experiment.

4 Analysis and Results

In our analysis, we will focus primarily on comparing the category construction (conceptualization) behavior of participants between the English, Chinese, and Korean experiments.

Our experimental software CatScan automatically recorded the number of groups created by each participant and the time (in seconds) the participant spent on the grouping task. On average, participants in the English experiment created 8.05 groups (SD 7.29) and spent 968.89 seconds on the grouping task (SD 824.43 seconds). In the Chinese experiment, participants created 7.80 groups (SD 6.81) on average and spent 777.97 seconds on the grouping task (SD 880.67 seconds). For the Korean experiment, participants created 5.30 groups (SD 2.45) on average and spent 617.30 seconds on the grouping task (SD 594.40 seconds). A one-way analysis of variance (ANOVA) comparing the number of categories created revealed that there are no statistically significant differences between English, Chinese, and Korean participants ($F(57,2) = 1.397$, $p = 0.260$). Similarly, there are no statistically significant differences in the time English, Chinese, and Korean participants spent on creating groups ($F(57,2) = 1.071$, $p = 0.350$).

The category construction behavior of each participant was recorded in an 88 * 88 sized individual similarity matrix (88 is the number of icons used in each experiment). In these matrices, the similarity of each pair of icons is binary-coded: A value of ‘1’ indicates that a pair of icons has been placed into the same group by that participant whereas a value of ‘0’ indicates that a pair of icons was not placed into the same group. An overall similarity matrix (OSM) is obtained by summing over the individual similarity matrices of all 20 participants in each experiment. Therefore, the similarity value of a pair of icons in an OSM ranges from 0 (lowest similarity possible) to 20 (highest similarity possible).

Figure 4 synthesizes several aspects of the different analyses we performed. The images present a combination of heat maps and dendrograms (generated by Ward’s clustering method) for the English, Chinese, and Korean experiments. The heat maps intuitively visualize the OSMs and, hence, the overall category construction behavior of all participants in each experiment. Red cells correspond to the highest similarity values (20) in the OSM, while white cells corresponds to the lowest similarity values (0) with the color varying between white and red on a continuous gradient. The dendrogram placed to the left side of each heat map was generated from a cluster analysis using Ward’s method. The order of the icons (both in rows and columns) were rearranged such that icons that are similar to each other in the cluster analysis always neighbor with one another.

Overlap relations in all three experiments (English, Chinese, and Korean) form three general clusters, corresponding to the three larger red blocks found along the diagonal of the heat maps (see Figure 4). The results of both analysis methods are corroborated by visually comparing the top three clusters with the matching placement of the three larger blocks (hot spots) found on the heat maps. Upon analyzing the clustering structure across all three heat map/dendrogram combinations in more detail, the three major categories can be summarized as follows:

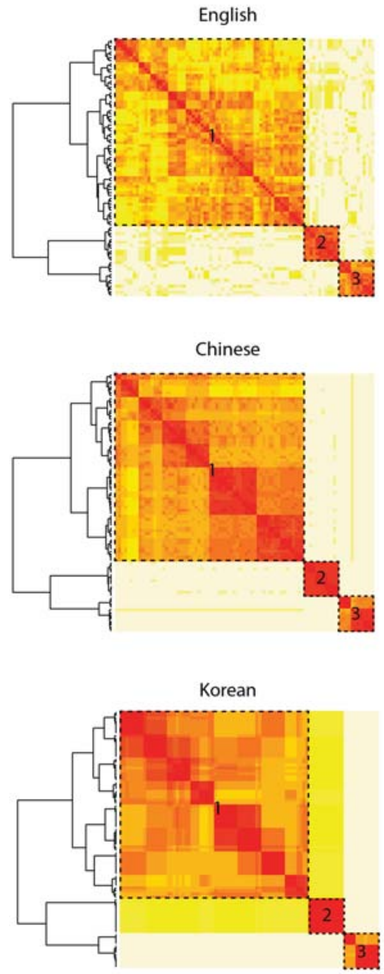

Fig. 4. Heat maps for the three experiments (English, Chinese, and Korean). Rows and columns are organized based on the results of Ward's clustering method as shown by the dendrograms. Numbers indicate a three cluster solution: 1. various overlap relations; 2. proper part relations; 3. non-overlapping relations.

- **Cluster 1** contains all `g`, `h`, `i`, `j`, `l`, `m`, `n`, `o`, `p`, `q`, `r`, `s`, `t`, `u`, `v`, and `w` icons. A commonality shared by all these icons (relations) is that the oil spill **overlaps** with the protected habitat zone.
- **Cluster 2** contains all `d`, `dd`, and `k` icons, in which the oil spill is found **inside** the protected habitat zones regardless of whether or not the boundary of the oil spill touches the boundary of the protected habitat zone.

– **Cluster 3** contains all `a`, `aa`, and `b` icons, in which the oil spill is found **outside** the protected habitat zones regardless of whether or not the boundary of the oil spill touches the boundary of protected habitat zone.

To validate this interpretation, we followed a suggestion by Clatworthy and collaborators [6] and compared the structures of the dendrograms generated using Average Linkage and Complete Linkage in addition to the dendrograms generated using Ward's method. The cross-examination shows that the three-cluster structure is identical across all three dendrograms in all three experiments and that it is the only cluster structure that is identical. Additionally, we randomly sampled subsets of 10 participants (three times) and each time performed cluster analyses on these subsets using the three different clustering algorithms. The three-cluster structure again is a) consistent and b) the only consistent clustering solution. This strongly supports the validity of the three cluster solutions and also indicates that all three language experiments are comparable.

Despite the fact that the three cluster solution can be considered the natural category structure (as revealed by the different validation procedures), differences exist within the internal structure of the overlap category across the three languages. Intrigued by the differences, we took a closer look at participants' grouping behavior specifically for the overlap relations from Cluster 1. We extracted a sub-matrix containing only similarity values (i.e., 0 or 1) for Cluster 1 relations from each individual similarity matrix. Based on these sub-matrices, we constructed a between-participant similarity matrix (BSM) that encodes the similarity of grouping behavior for each pair of participants for each of the three experiments. In the BSM, the similarity value of each pair of participants is determined by the Hamming distance between the sub-matrices of these two participants. Hamming distance calculates the number of cells that differ across two matrices, thus effectively calculating the distance, or difference, between two participants' grouping behavior for Cluster 1 relations. Furthermore, a dendrogram was generated using Ward's method for each experiment, showing the similarity of participants based on their grouping behavior on Cluster 1 relations (see Figure 5 as an example from the Chinese experiment).

Guided by the results in Figure 5, we explored the groups created by participants and their linguistic description in KlipArt—a custom-made, freely available visual analysis tool [31]. In KlipArt, we were able to identify five major types of category construction strategies (as specified in Table 2) employed by participants in all three experiments.

Overall, more Chinese and Korean participants than English participants placed all overlap relations into one single group while more English participants grouped icons by the direction the oil spill faced than the Chinese and Korean participants. Korean participants grouped mainly based on the penetration of the oil spill in the protected habitat zone whereas the other two participant groups were more spread out among category construction strategies. For the final two grouping strategies, a similar number of participants (comparing between the Chinese, English, and Korean participants) were found to behave and group in the same way (i.e. grouping by the amount of the overlap: 20-25%

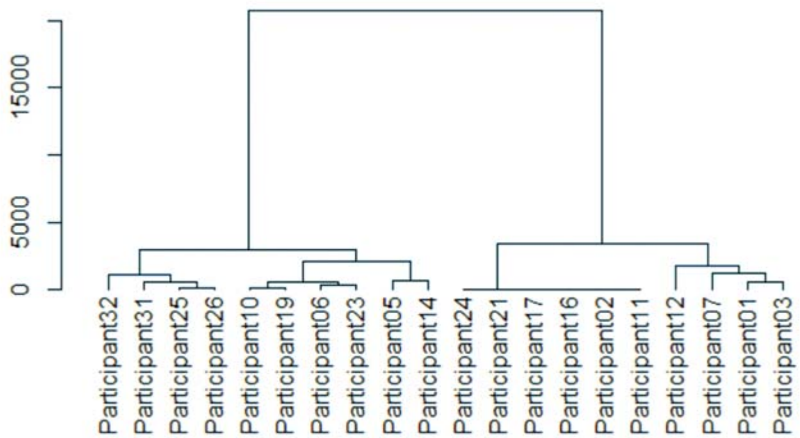

Fig. 5. This dendrogram shows the similarities of participants from the Chinese experiment based on their grouping behavior on Cluster 1 relations. The lower the distance at which two participants are merged with each other, the more similar they are.

Table 2. Category constructions strategies elicited from participants (P) for the different languages

Grouping strategy	English experiment	Chinese experiment	Korean experiment
All overlap icons placed in one group	3 (15%)	6 (30%)	7 (35%)
Grouping by the amount of the overlap (large vs. small)	4 (20%)	4 (20%)	5 (25%)
Topological equivalence classes (according to modes of overlap relations)	2 (10%)	2 (10%)	0 (0%)
Whether the long or short tail of the oil spill penetrates the protected habitat zone or not	5 (25%)	6 (30%)	8 (40%)
Direction that the oil spill is facing	6 (30%)	2 (10%)	0 (0%)

of participants; grouping by the topological equivalence classes implied by the modes of overlap relations: 0-10% of participants). This approach allows us to undergo finer-grained analyses that compare potential language related differences and grouping strategies. Furthermore, these findings allow us to identify new research questions to ascertain why there were grouping differences between the experimental groups. For instance, future research can analyze why English participants utilize direction more than participants in the other two language groups.

5 Conclusions

5.1 General discussion

The results shed light on a number of essential details about humans' conceptualization of modes of overlap across different languages. Galton (1998) is making a convincing case that it would be desirable to adopt a finer level of granularity with respect to distinguishing modes of overlap rather than one overlap relation (as in RCC and IM). In contrast, our results indicate that this level of granularity is not common from a cognitive perspective in either of the three languages. The cross validation techniques applied to all three data sets reveal an extremely stable category structure distinguishing non-overlapping, overlapping, and proper part relations. This coarse and consistent level of granularity is remarkable for a number of reasons: First, we generally find that if spatial information is statically presented, distinctions made by qualitative calculi are reasonable predictors for category construction behaviors. This has been shown, for example, by Knauff and collaborators [33]. They revealed in their experiments using the eight relations identified by RCC/IM that participants did abide by formal distinctions. In contrast, in the results presented here, a very coarse perspective on spatial relations is adopted, that is, 22 different relations form only three spatial categories. Second, none of the languages shows a diversion from this clustering structure although previous research indicated that native Korean speakers might be inclined to distinguish different proper part relations [51]. Third, while finer distinctions are adopted by participants in all three experiments, they do not surface as a consistent pattern. Hence, they should not be considered as being fundamental (or salient discontinuities). This interpretation remains true even if participants are explicitly asked to create as many meaningful categories as possible [69].

5.2 Reflections on research framework

Our ultimate goal is the development of a research framework that allows for comprehensive behavioral research on fundamental concepts of space and time and space-time, Func2st. It is widely acknowledged (e.g., [32, 35]) that equivalence classes identified by qualitative spatio-temporal calculi offer testable hypotheses that take the following general form: *Formally identified equivalence classes correspond to salient, quasi-symbolic equivalence that surfaces in reliable cognitive behavior*. In other words, discontinuities of space, time, and space-time relevant for cognitive information processing correspond to discrete distinctions introduced by qualitative spatio-temporal calculi. Our research framework adds an important aspect to the existing literature. In contrast to research that starts out analyzing language components such as spatial and temporal preposition, our work starts with eliciting conceptualizations of spatial relations. Given the often controversial discussion of the relation between language and cognitive concepts that ranges from assuming universal conceptual categories that are essentially uniform across languages [57, 41, 67] to assuming language dependent concepts [40] via recent approaches that propose a mapping between concepts and words in form of word-clusters [22], we consider it essential to add a non-linguistic approach that might be able to shed some light on the conceptual structure of human concepts and whether or not they are universal (even if no extensive linguistic theory exists).

5.3 Reflections on crowdsourcing

Crowdsourcing is a powerful tool to scale up the behavioral calibration of formally (qualitatively) identified discontinuities of space and time. This approach is necessary as there is a large number of potential candidates for fundamental spatial concepts that require an efficient approach for validation.

Yet, there are some challenges in using the crowd that are important to keep in mind. When it comes to testing concepts using English speaking participants crowdsourcing is indeed a powerful and efficient tool for behavioral research. AMT, the platform we used for our experiments, allowed us to set very high standards with respect to Turker reliability and, nonetheless, data came in overnight (for English HITs). While we needed some experimentation with the exact phrasing of the experimental task and the reward structure, we have additional evidence in form of a pre-study, that the results obtained through AMT are comparable to regular lab studies. Our results are therefore in line with many current studies (e.g., [59]) that show that AMT is a reliable research tool. Additionally, and maybe even more importantly, using AMT can overcome the critical aspect that more lab studies rely almost exclusively on student participants. Several studies have shown that the population of Turkers is much more diverse than any undergraduate student body (e.g., [50]).

With respect to using AMT for cross cultural studies we have obtained mixed results. The Chinese version of our experiment allowed us to have similarly high standards as the English version. However, the data came in at a much slower pace and it took us roughly two weeks to collect data for 20 participants. We did not require participants to live in China but to be native language speakers of Chinese (compare [53]). The long time it took makes this approach difficult in case larger numbers are needed but it seems possible to use AMT for Chinese studies. Additionally, it may be possible to increase participant numbers by using higher payments and/or lower Turker ratings. One option we did not fully explore yet is advertising HITs outside of AMT.

The Korean version of the experiments was extremely slow such that we collected additional data through direct requests. We experimented with higher payments and lower Turker ratings but got mixed results. More Turkers signed up for our HITs but we also had to disregard data as we obtained invalid submissions. The most extreme case was a participant whose linguistic descriptions in the second part were, most likely, obtained through an online translation service such as Google Translate.

6 Outlook

We only provided a glimpse of how fundamental cognitive spatial, temporal, and spatio-temporal concepts, Func2st, can be evaluated in a crowdsourcing, cross-linguistic setting. We selected Galton's modes of overlap as a test case for our quest to deliver a comprehensive assessment of Func2st relating formal and human conceptualizations of space and time. However, the research framework we have developed is tailored to be efficient and effective such that a large scale assessment is indeed becoming feasible. We believe that every formalism that is proposed to enhance processes at the interface of humans and machines requires some validation. We have provided an overview of potential extensions of our work in Table 1. By no means do we claim that this list is complete and we invite others to comment and contribute to this ongoing research challenge.

References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM 26(11), 832–843 (1983)
2. Aziz-Zadeh, L., Casasanto, D., Feldman, J., Saxe, R., Talmy, L.: Discovering the conceptual primitives. In: Love, B., McRae, K., Sloutsky, V.M. (eds.) Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 27–28. Cognitive Science Society, Austin and TX (2008)
3. Bloom, P., Peterson, M.P., Nadel, L., Garrett, M.F. (eds.): Language and space. MIT Press, Cambridge and MA (1996)
4. Boroditsky, L.: Does language shape thought?: Mandarin and English speakers' conceptions of time. Cognitive Psychology 43, 1–22 (2001)
5. Brunet, R.: La carte, mode d'emploi. Fayard-Reclus, Paris (1987)
6. Clatworthy, J., Buick, D., Hankins, M., Weinman, J., Horne, R.: The use and reporting of cluster analysis in health psychology: A review. British Journal of Health Psychology 10, 329–358 (2005)
7. Cohn, A.G., Renz, J.: Qualitative spatial representation and reasoning. In: van Harmelen, F., Lifschitz, V., Porter, B. (eds.) Handbook of knowledge representation, pp. 551–596. Foundations of artificial intelligence, Elsevier (2008)
8. Corcoran, P., Mooney, P., Bertolotto, M.: Spatial Relations Using High Level Concepts. ISPRS International Journal of Geo-Information 1(3), 333–350 (2012)
9. Crawford, L.E., Regier, T., Huttenlocher, J.: Linguistic and non- linguistic spatial categorization. Cognition 75(3), 209–235 (2000)
10. Egenhofer, M.J.: The family of conceptual neighborhood graphs for region-region relations. In: Fabrikant, S.I., Reichenbacher, T., van Krefeld, M., Schlieder, C. (eds.) Sixth International Conference, GIScience 2010 Zürich, Switzerland, Sep. 14-17 2010, Proceedings, pp. 42–55. Springer, Berlin (2010)
11. Egenhofer, M.J., Franzosa, R.D.: Point-set topological spatial relations. International Journal of Geographical Information Systems 5(2), 161–174 (1991)
12. Egenhofer, M.J., Mark, D.M.: Naive geography. In: Frank, A.U., Kuhn, W. (eds.) Spatial Information Theory, pp. 1–15. Springer, Berlin (1995)
13. Galton, A.: Modes of overlap. Journal of Visual Languages and Computing 9, 61–79 (1998)

14. Galton, A.: Qualitative spatial change. Spatial information systems, Oxford Univ. Press, Oxford (2000)
15. Gibson, J.: The ecological approach to visual perception. Houghton Mifflin (1979)
16. Goldstone, R.L., Hendrickson, A.T.: Categorical perception. Wiley Interdisciplinary Reviews: Cognitive Science 1(1), 69–78 (2010)
17. Gumperz, J.J., Levinson, S.C. (eds.): Rethinking linguistic relativity. Cambridge University Press, Cambridge (1996)
18. Harnad, S.: To cognize is to categorize: Cognition is categorization. In: Cohen, H., Lefebvre, C. (eds.) Handbook of categorization in cognitive science. Elsevier, Amsterdam (2005)
19. Hayes, P.: The naive physics manifesto. In: Michie, D. (ed.) Expert Systems in the Microelectronic Age, pp. 242–270. Edinburgh University Press, Edinburgh (1978)
20. Heer, J., Bostock, M.: Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In: ACM CHI 2010 (2010)
21. Hermer-Vazquez, L., Moffet, A., Munkholm, P.: Language, space, and the development of cognitive flexibility in humans: the case of two spatial memory tasks. Cognition 79, 263–299 (2001)
22. Holmes, K.J.: Language as a Window into the Mind: The Case of Space. Ph.D. thesis, Emory University (2012)
23. January, D., Kako, E.: Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). Cognition 104, 417–426 (2007)
24. Johnson, M.: The body in the mind: The bodily basis of meaning, imagination, and reasoning. University of Chicago Press, Chicago (1987)
25. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. pp. 453–456. ACM (2008)
26. Klein, F.: Vergleichende Betrachtungen über neuere geometrische Forschungen ("A comparative review of recent researches in geometry"). Mathematische Annalen 43, 63–100 (1872)
27. Klippel, A.: Spatial information theory meets spatial thinking - Is topology the Rosetta Stone of spatio-temporal cognition? Annals of the Association of American Geographers 102(6), 1310–1328 (2012)
28. Klippel, A., Hardisty, F., Li, R.: Interpreting spatial patterns: An inquiry into formal and cognitive aspects of Tobler's first law of geography. Annals of the Association of American Geographers 101(5), 1011–1031 (2011)
29. Klippel, A., Montello, D.R.: Linguistic and nonlinguistic turn direction concepts. In: Winter, S., Kuipers, B., Duckham, M., Kulik, L. (eds.) Spatial Information Theory, pp. 354–372. Springer, Berlin (2007)
30. Klippel, A., Wallgrün, J.O., Yang, J., Li, R., Dylla, F.: Formally grounding spatio-temporal thinking. Cognitive Processing 13(Supplement 1), 209–214 (2012)
31. Klippel, A., Weaver, C., Robinson, A.C.: Analyzing cognitive conceptualizations using interactive visual environments. Cartography and Geographic Information Science 38(1), 52–68 (2011)
32. Klix, F.: Die Natur des Verstandes. Hogrefe, Göttingen (1992)
33. Knauff, M., Rauh, R., Renz, J.: A cognitive assessment of topological spatial relations: Results from an empirical investigation. In: Hirtle, S.C., Frank, A.U. (eds.) Spatial information theory: A theoretical basis for GIS, pp. 193–206. Springer, Berlin (1997)
34. Knauff, M., Ragni, M.: Cross-Cultural Preferences in Spatial Reasoning. Journal of Cognition and Culture 11(1), 1–21 (2011)

35. Kuhn, W.: An image-schematic account of spatial categories. In: Winter, S., Kuipers, B., Duckham, M., Kulik, L. (eds.) Spatial Information Theory, pp. 152–168. Springer, Berlin (2007)
36. Kurata, Y., Egenhofer, M.J.: Interpretation of behaviors from a viewpoint of topology. In: Gottfried, B., Aghajan, H. (eds.) Behaviour monitoring and interpretation, pp. 75–97. IOS Press, Amsterdam (2009)
37. Lakoff, G.: Women, fire and dangerous things. Chicago University Press (1987)
38. Lakoff, G.: The invariance hypothesis: is abstract reason based on image schemata? Cognitive Linguistics 1(1), 39–74 (1990)
39. Levinson, S.C., Kita, S., Haun, D., Rasch, B.H.: Returning the tables: language affects spatial reasoning. Cognition 84, 155–188 (2002)
40. Levinson, S.C., Meira, S., The Language and Cognition Group: 'Natural concepts' in the spatial topological domain—Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. Language 79(3), 485–516 (2003)
41. Li, P., Gleitman, L.: Turning the tables: Language and spatial reasoning. Cognition 83, 265–294 (2002)
42. Ligozat, G.: Qualitative spatial and temporal reasoning. ISTE, London (2012)
43. Lu, S., Harter, D., Graesser, A.C.: An empirical and computational investigation of perceiving and remembering event temporal relations. Cognitive Science 33, 345–373 (2009)
44. Maguire, M.J., Brumberg, J., Ennis, M., Shipley, T.F.: Similarities in object and event segmentation: A geometric approach to event path segmentation. Spatial Cognition and Computation (3), 254–279 (2011)
45. Malt, B.C., Ameel, E., Gennari, S., Imai, M., Saji, N., Majid, A.: Do words reveal concepts? In: Carlson, L.A., Hölscher, C., Shipley, T.F. (eds.) Proceedings of the 33rd Annual Conference of the Cognitive Science Society., pp. 519–525. Cognitive Science Society, Austin and TX (2011)
46. Mandler, J.M.: How to build a baby II. Conceptual primitives. Psychological Review 99(4), 587–604 (1992)
47. Mandler, J.M.: The spatial foundations of the conceptual system. Language and Cognition 2(1) (2010)
48. Mark, D.M., Egenhofer, M.J.: Calibrating the meanings of spatial predicates from natural language: Line-region relations. In: Waugh, T.C., Healey, R.G. (eds.) Advances in GIS Research, 6th International Symposium on Spatial Data Handling, Edinburgh, Scotland, UK., pp. 538–553 (1994)
49. Mark, D.M., Egenhofer, M.J.: Topology of prototypical spatial relations between lines and regions in English and Spanish. In: Proceedings, Auto Carto 12, Charlotte, North Carolina, March 1995,, pp. 245–254 (1995)
50. Mason, W.A., Suri, S.: Conducting behavioral research on amazon's mechanical turk. Behavior Research Methods, Instruments, & Computers 44(1), 1–23 (2011)
51. McDonough, L., Choi, S., Mandler, J.M.: Understanding spatial relations: Flexible infants, lexical adults. Cognitive Psychology 46(3), 229–259 (2003)
52. Medin, D.L., Wattenmaker, W.D., Hampson, S.E.: Family resemblance, conceptual cohesiveness, and category construction. Cognitive Psychology 19(2), 242–279 (1987)
53. Munnich, E., Landau, B., Dosher, B.A.: Spatial language and spatial representation: A cross-linguistic comparison. Cognition 81, 171–207 (2001)
54. Op Beeck, H.P.d., Wagemans, J., Vogels, R.: The representation of perceived shape similarity and its role for category learning in monkeys: A modeling study. Vision Research 48(4), 598–610 (2008)

55. Paolacci, G., Chandler, J., Ipeirotis, P.G.: Running experiments on amazon mechanical turk. Judgement and Decision Making 5(5), 411–419 (2010)
56. Peuquet, D.J.: Representations of space and time. Guilford Press, New York (2002)
57. Pinker, S.: The language instinct: How the mind creates language. HarperPerennial ModernClassics, New York, 1 edn. (2007)
58. Pothos, E.M., Chater, N.: A simplicity principle in unsupervised human categorization. Cognitive Science 26(3), 303–343 (2002)
59. Rand, D.G.: The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments: Evolution of Cooperation. Journal of Theoretical Biology 299(0), 172–179 (2012)
60. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connections. In: Nebel, B., Rich, C., Swartout, W.R. (eds.) Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning, pp. 165–176. Morgan Kaufmann, San Francisco (1992)
61. Renolen, A.: Modelling the Real World: Conceptual Modelling in Spatiotemporal Information System Design. Transactions in GIS 4(1), 23–42 (2000)
62. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers? Shifting demographics in mechanical turk. In: CHI 2012: Imagine all the People, pp. 2863–2872 (2012)
63. Scheider, S., Janowicz, K., Kuhn, W.: Grounding geographic categories in the meaningful environment. In: Stewart Hornsby, K., Claramunt, C., Denis, M., Ligozat, G. (eds.) Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings. Springer, Berlin (2009)
64. Scheider, S., Kuhn, W.: Affordance-based categorization of road network data using a grounded theory of channel networks. International Journal of Geographical Information Science 24(8), 1249–1267 (2010)
65. Shaw, R., McIntyre, M., Mace, W.: The role of symmetry in event perception. In: MacLeod, R.B., Pick, H.L. (eds.) Perception, pp. 276–310. Cornell University Press, Ithaca (1974)
66. Smith, B., Mark, D.M.: Geographical categories: An ontological investigation. International Journal of Geographical Information Science 15(7), 591–612 (2001)
67. Stock, K., Cialone, C.: Universality, Language-Variability and Individuality: Defining Linguistic Building Blocks for Spatial Relations. In: Egenhofer, M., Giudice, N., Moratz, R., Worboys, M. (eds.) Spatial Information Theory., vol. 6899, pp. 391–412. Springer, Berlin (2011)
68. Talmy, L.: How language structures space. In: Pick, H.L., Acredolo, L.P. (eds.) Spatial orientation: Theory, research, and application, pp. 225–282. New York (1983)
69. Wallgrün, J.O., Yang, J., Klippel, A.: Investigating intuitive granularities of overlap relations. In: Proceedings of the 12th IEEE International Conference on Cognitive Informatics & Cognitive Computing, New York City, USA. Wiley-IEEE CS Press, Hoboken, NJ (2013)
70. Wierzbicka, A.: The search for universal semantic primitives. In: Pütz, M. (ed.) Thirty years of linguistic evolution, pp. 215–242. Benjamins (1992)
71. Wolff, P., Holmes, K.J.: Linguistic relativity. Wiley Interdisciplinary Reviews: Cognitive Science 2(3), 253–265 (2011)
72. Worboys, M., Duckham, M.: Monitoring qualitative spatiotemporal change for geosensor networks. International Journal of Geographical Information Science 20(10), 1087–1108 (2006)
73. Worboys, M., Duckham, M.: GIS: A Computing Perspective. CRC Press, Boca Raton FL, 2 edn. (2004)