REVIEW ARTICLE

Quantifying space, understanding minds: A visual summary approach

Mark Simpson[1], Kai-Florian Richter[2], Jan Oliver Wallgrün[1], and Alexander Klippel[1]

[1]The Department of Geography, The Pennsylvania State University, PA, USA.
[2]Department of Geography, University of Zurich, Zurich, Switzerland.

Abstract: This paper presents an illustrated, validated taxonomy of research that compares spatial measures to human behavior. Spatial measures quantify the spatial characteristics of environments, such as the centrality of intersections in a street network or the accessibility of a room in a building from all the other rooms. While spatial measures have been of interest to spatial sciences, they are also of importance in the behavioral sciences for use in modeling human behavior. A high correlation between values for spatial measures and specific behaviors can provide insights into an environment's legibility, and contribute to a deeper understanding of human spatial cognition. Research in this area takes place in several domains, which makes a full understanding of existing literature difficult. To address this challenge, we adopt a visual summary approach. Literature is analyzed, and recurring topics are identified and validated with independent inter-rater agreement tasks in order to create a robust taxonomy for spatial measures and human behavior. The taxonomy is then illustrated with a visual representation that allows for at-a-glance visual access to the content of individual research papers in a corpus. A public web interface has been created that allows interested researchers to add to the database and create visual summaries for their research papers using our taxonomy.

Keywords: space syntax, spatial measures, environmental cognition, visual summaries

1 Introduction

What makes some places easy to get lost in, but others eminently navigable? How does the way streets are laid out in a city affect where people go within it? Answering these and similar questions requires models that relate the spatial characteristics of an environment to human spatial behavior within that environment [16, 26]. The link between spatial

characteristics and behavior has been well established, but most findings rely on qualitative descriptors that make comparing findings difficult [65]. Comparing behavior to spatial characteristics more systematically requires the formalization of those characteristics, which we refer to here as *spatial measures*. Broadly, spatial measures are techniques and methods for quantitatively measuring space and spatial relationships. Importantly, they can move beyond classical metric understanding (e.g., distance and direction) and measure aspects such as relative centrality. They arise in several areas of research, perhaps most prominently within the realm of *space syntax* [27]. They allow for different environments to be directly compared in terms of their spatial characteristics, and to quantitatively compare those characteristics to various spatial behavioral measures (e.g., wayfinding performance).

The primary motivation for understanding the relationship between spatial measures and human behavior is to gain insight into the underlying mental processes that drive human behavior in space. Spatial measures can also be used as behavioral models directly (i.e., to predict behavior in an environment) in fields such as architecture and urban planning [16, 24]. Understanding how spatial characteristics of an environment may influence human behavior is of wider importance because it can provide insights into how to compensate for difficulties in understanding an environment, such as during wayfinding, and into how to create spaces that are more easily navigable. Almost everyone has become lost in a confusing building at one time or another, and some places seem difficult to navigate even if one has been there many times before [11]. What may be a mere inconvenience in normal activity may become critical in emergency situations.

Despite the importance of this research domain, an overview of existing research to help synthesize knowledge across disciplinary boundaries is missing (see Bafna [2], Carlson et al. [11], and Dalton et al. [15] for existing reviews in the area). We use the visual summary approach, adapted from Mason et al. [41] to create an illustrated taxonomy that is intended to enable such an overview. This method involves the manual review of selected literature to construct taxonomic categories, which are represented in a diagram called a visual summary. Literature is selected according to its relevance as judged by the authors, with the intention of capturing the breadth of the target research area. Once a paper is classified, a visual summary can be made to show which categories it contains. A series of visual summaries then allows a user to quickly compare a series of papers according to their content (see Figure 5 for an example). We have implemented a web interface which allows authors (or other interested parties) to classify literature and create visual summaries themselves, which consequently will grow the database of classified literature and, over time, will allow for increasingly detailed analyses.

The remainder of the paper is structured as follows: Section 2 provides background on research that measures space in order to predict human behaviors and discusses the need for a novel taxonomy, and our selection of the visual summary method. In Section 3, we cover the methods for creating the taxonomy and visual summary. Section 4 presents the final visual summary design and the definitions for the categories illustrated in the visual summaries, discusses categories that were considered but not included, and then reports the results of our inter-rater validation tasks. We then give an example of a visual summary applied to a single piece of literature, and a user-centric example of how multiple visual summaries can be compared and research patterns identified. We also briefly describe the open-access web interface that allows users to browse the existing database and create their own visual summaries. In Section 5 we provide an outlook on potential extensions of the method and use of the database of literature.

2 Background

In this section, we will review key trends in environmental cognition and spatial measures research in order to provide some context to the taxonomy and visual summary we have created. This section is kept deliberately short, as a significant amount of research is reviewed in the course of explaining the taxonomy itself, in Section 4.1.

2.1 Environmental cognition and spatial measures

A fundamental challenge in this area of research is that human spatial memory structures are distorted, and that the degree and type of distortion can vary with the spatial properties of an environment [17, 67]. Following Golledge [22], we refer to these spatial memory structures as *internal representations*. Since the biological basis for spatial information storage is not thoroughly understood in humans, "internal representations must be inferred from one or more external symbolic representations (e.g., sketch maps of a city) or from some other forms of observable behavior (e.g., search behavior to find a specific location)" [22]. Direct observation of internal representations is currently not possible, so researchers must instead examine behaviors which make use of them.

Spatial measures allow for more objective connections between spatial behavior and the spatial aspects of the environment. The basic premise is: if a measure correlates well with some spatial behavior within an environment, there must be some connection between how the measure represents (or operates on) that environment and how the human brain does. However, this relationship is not easy to establish. There are many potential stimuli involved in creating environmental knowledge, and individual strategies for activities such as wayfinding differ considerably [66]. This indicates that it may be necessary to combine different measures to capture the variety and combinations of spatial aspects used by humans in order to fully understand how internal representations are created and used [20].

2.2 Spatial measures

The research in this paper is intended to bridge knowledge from several disciplines, but is most closely related to the field of environmental cognition [17], which traces its origins to Lynch's Image of the City [39]. Lynch provided much of the impetus for research into the relationship between human understanding of an environment and the spatial aspects of the environment itself. He proposed a deep connection between the structure of an environment—in particular urban areas—and a person's ability to understand that environment, showing that some structures make places inherently more understandable. He coined the term *environmental legibility*, which refers to the ease with which an environment can be transformed into a series of *mental images* (i.e., internal representations). Lynch identified nodes, paths, edges, districts, and landmarks as the salient elements for environmental legibility and, consequently, for internal representations themselves. Importantly, Lynch's descriptors were qualitative, and therefore not true spatial measures as considered in this research. They were derived from interviews with people who regularly encountered those spaces (urban centers) and from sketch maps they made rather than direct measurement of the environment. However, Lynch's elements were an important

early attempt to systematically understand how environmental aspects affect internal representations.

Space syntax is perhaps the most prevalent research program within the domain of spatial measures and human behavior research. Space syntax originated with Hillier and Hanson [27], who were concerned with how societies could be understood through how people arranged inhabited spaces, such as rooms within a building, or streets within an urban area [2]. In order to accomplish this, they created systematized abstraction techniques that represent spaces with graphs, where sub-spaces are identified by nodes and connected spaces linked in the graph. This allows the use of topologically-derived measures, such as the degree of a node.

Space syntax has generated many methods for abstracting environments into graphs and explored topological measures of those graphs, and compared them to a variety of human behaviors. One of the earliest and most common abstractions is the axial map, which breaks environments into inter-visible chunks represented by axial lines [27]. Axial lines are the longest unbroken lines that can be drawn from one part of a space through another, beginning with the longest possible line in an environment and continuing until all spaces are accounted for within that environment. Axial line maps are often quantified with a novel graph measure called integration, and compared to pedestrian traffic [27]. Integration is a form of network centrality, used for example in [33, 50], and is often coupled with connectivity (degree of a node), as in [37, 47].

Other methods pioneered by the space syntax community include visibility graph analysis (VGA) [62] and segment analysis [61], as well as extensions to axial line measurement, such as angular analysis [59] and extended axial lines [1]. The space syntax community has likewise applied measures to a wide variety of human behaviors. For example, Asami et al. [1] compared these measures to the number of stories buildings had to identify "local centers" in historic Istanbul, and Baran et al. [3] compared walking behavior in neighborhoods with different types of street networks identified through applying space syntax measures. Such works reflect the traditional focus of space syntax on sociological understanding, but space syntax researchers have also branched out into explicitly cognitive research.

According to Bafna [2], space syntax moved towards spatial cognition because it "has always sought to examine the relationship between behavior and space by examining behavior not merely with respect to its local setting (in which the perceptual typically dominates) but with respect to the global setting in which it occurs, where the cognitive dimension of behavior comes into play by necessity." Dalton and Hölscher [15] discuss space syntax and cognition research in detail, and point to Peponis et al. [51] as the first synthesis of space syntax and cognitively-aware research. Notably, Montello [44] has critiqued space syntax's approach to cognitive problems, but nevertheless praised its generation of methods for quantifying space. Perhaps unsurprisingly, spatial measures derived from the space syntax community continue to be used in cognitively-focused research, such as Hölscher et al. [29] and Li and Klippel [37].

Quantitative spatial measures that are applied to human behavioral problems are relatively rare outside of the broader space syntax community. Two notable examples are isovist analysis and interconnection density (ICD). ICD was developed by O'Neill [49], who compared different environments quantified with ICD to participant performance in those environments. Li and Klippel [37], built upon O'Neill's work by combining ICD with space syntax analyses.

Isovists were invented by Hardy [25], but named as such by Tandy [57]. It was Benedikt [5], however, who created and popularized methods for quantifying their attributes, as well as extending the concept to isovist fields. Isovists measure the properties of visible space across an environment. A single point isovist is a polygon that represents the viewable area around a single point in an environment, essentially equivalent to a viewshed in a geographic information science context [63]. In order to account for the more natural limited perspective of humans, partial isovists with restricted viewing angles have also been used [42]. Since isovists are treated as polygons, they can be quantified with attributes such as perimeter, area, number of points (corners), and ratios of those attributes to one another. Successful applications of isovist measures in environmental cognition problems include Wiener and Franz [65], and Meilinger, Franz, and Bulthoff [42]. The former had participants navigate to the point in a room with maximum visibility and rate experiential properties of spaces. In the latter, participants navigated a virtual city and had their route and landmark knowledge tested. Isovists also inspired attempts to assimilate ideas about how to quantify visibility with space syntax measures, most notably in visibility graph analysis [62].

2.3 The need for a taxonomy

While a considerable amount of work has been done in different disciplines along the lines of comparing behavior to quantifiable aspects of space [65], few attempts have been made to systematically understand if and how approaches differ [16]. Zimring and Dalton [67] note that despite progress made in environmental cognition, "as with many multidisciplinary fields, however, communication among researchers is uneven." While discussing the problem of relating environmental layouts to human behavior, Franz and Wiener [20] state that a unifying framework for measurement methods is needed to account for the complexity of the real world. More practically, combinations of spatial measures promise to capture more salient aspects of the environment at once, but the plethora of existing measures makes the path forward uncertain. Despite the call for one within the literature, a comprehensive understanding of the progress accomplished in this already complex area has remained elusive.

2.4 Visual summaries

The current lack of a big-picture view when it comes to spatial measures and human behavior research indicates the need for a taxonomy, but how to go about creating and spreading it to a wider audience? On both accounts we adopt the approach of Mason et al. [41]. For a given field of literature, a taxonomy is systematically created, validated (see Section 4.3), and illustrated (see Section 4.8). The first result is a robust taxonomy that provides context for the target field. The second result is a design for a visual summary diagram that serves as a template for summary diagrams for individual pieces of literature. These diagrams illustrate the content of a paper (according to the taxonomy), but also display which categories it does *not* contain. This allows users to easily compare many pieces of literature by using multiple summary diagrams, while remaining cognizant of the entire taxonomy.

These diagrams make the literature content more accessible by giving users carefully designed *visual* access to it. This is intended to leverage the advantages of visual information, such as enabling quicker information processing and better pattern recognition

compared to textual information (see Mason et al. for a more detailed discussion [41]). The power of visualizing information instead of displaying it in textual or numerical form is well known, as explained by Bertin [6], Tufte [59], and Thomas and Cook [58], among many others. Particularly relevant in this context is Larkin and Simon's [36] work, which elaborates in detail on "Why a diagram is (sometimes) worth ten thousand words." But as Mason et al. [41] point out, despite such work exhorting the ability of visualization to assist in information comprehension, relatively few visual approaches have been taken to summarizing research literature.

Work in knowledge domain visualization is a possible exception, as it is also aimed at giving a visual overview of areas' research by analyzing literature [7]. However, as far as we are aware these approaches differ greatly from the visual summary method we adopt (see Section 4.7 for a discussion of our method compared to automated approaches). Domain visualization often focuses on understanding the relationships between groups of literature, authors of that literature, and larger topic areas (domains). In contrast, we emphasize the importance of portraying the content and concepts contained within *individual* pieces of literature in addition to giving an overview of a research area. Domain visualization often uses automated techniques to both analyze text and bibliometrically identifiable information such as authorships and citations to construct an overview of comparatively large corpora [7]. For example, Boyack et al. [9] uses over a million journal articles to construct a visualization of all sciences. Borrett et al. [8] whose focus was network ecology, had a somewhat more modest corpus of 33,900 articles.

The use of self-organizing maps (SOMs) [32, 55] is popular in domain visualization. A recent example can be seen in Skupin et al. [56], in which over two million publications are used to create a "map" of medical knowledge. In contrast, the affinity diagramming and visual summary method we adopt uses manual review to yield a taxonomy of a highly focused literature set, where the advantages of automation are more limited, and their application could even be counterproductive (for a more detailed discussion, see Section 4.7). For example, differing terminology among disciplines (e.g., *graph* versus *network*) can pose challenges for automated text analysis methods.

3 Methods

In this section, we discuss our method as it has been adapted from Mason et al. [41]. This includes how categories are generated using a modified affinity diagramming method, and the construction of a visual summary.

3.1 Creation of hierarchical categories

To create the categorization, we utilize a modified *affinity diagramming method*, adapted from Mason et al. [41] (see Figure 1), which is in turn based on the affinity diagramming method used by Skeels et al. [54]. Affinity diagramming identifies a set of categories that capture important themes and topics within a selected area of research by having experts place concepts into groups. Mason et al. extended this approach to be iterative and based on continuous literature evaluation, which we take farther with the addition of a naïve rater.

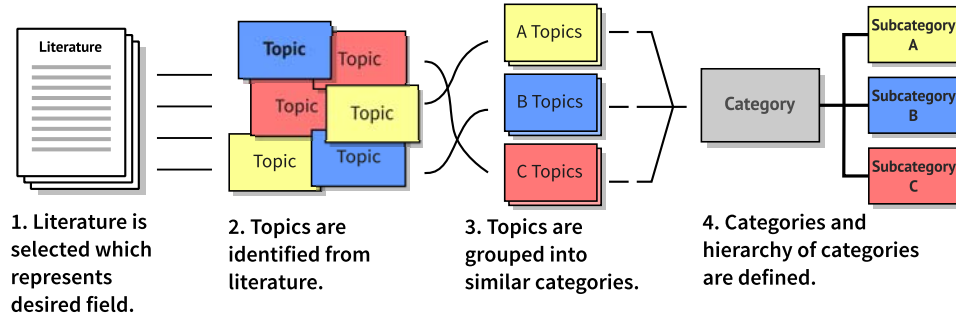## Modified Affinity Diagramming

Figure 1: Modified Affinity Diagramming, adapted from Mason et al. [41].

The researchers review a corpus of literature within the field of interest, create topics for the most salient aspects or themes, group them into categories, and create formal definitions for those categories. As more literature is read, categories are iteratively refined and structured into hierarchies, until a classification hierarchy deemed comprehensive is created. The comprehensiveness is evaluated via a classification task, in which researchers assign categories to a selection of papers without consulting each other, then measure their inter-rater agreement with Cohen's Kappa [13]. In our adaptation, we use multiple classification tasks between authors with an additional classification task with a naïve rater for an outside perspective. The specific results of our tasks are discussed below.

3.2 The visual summary diagrams

Visual summary diagrams are used to illustrate the content of a piece of literature according to our taxonomy, and idea adopted from Mason et al. [41]. Within the diagram, visual elements (arranged colored shapes) represent categories, and the overall design of the visual summary illustrates the structure of the categories found for a field (see Section 4.8 for a discussion on the design choices underlying the final diagram). Categories present within the paper have the corresponding visual element highlighted, while those not present are made less visible (but still legible, to give context of what was not contained within a paper). Thus, a new visual summary diagram is created for each piece of literature reviewed (see Appendix A: Additional Visual Summaries for examples) which shows the subjects (categories) covered within the specific text. This allows for an at-a-glance evaluation of how individual papers approach the field. Figure 2 shows our resultant visual summary template for the entire field of spatial measures and behavior (i.e., the categories identified within existing literature).

4 Results and discussion

In this section, we present our visual summary (Figure 2), which represents the categories contained in the surveyed literature, as well as the definitions of those categories. Additionally, we discuss some of the categories which were considered and why they were not

included, which includes topics we expected to find but that no current literature appears to address. We also build upon Mason et al.'s method by increasing the use of validation tasks, including using a naïve rater, to ensure that the categories were not arbitrary or idiosyncratic to the authors. We will also discuss the visual design of the visual summary diagram itself. We walk through an example visual summary for O'Neill [49] to give a practical example of how categories are assigned. We then discuss our implementation of a web interface that allows users to create their own visual summaries and contribute to the corpus. Finally, two example use cases are discussed to illustrate how visual summaries can be used for research tasks.
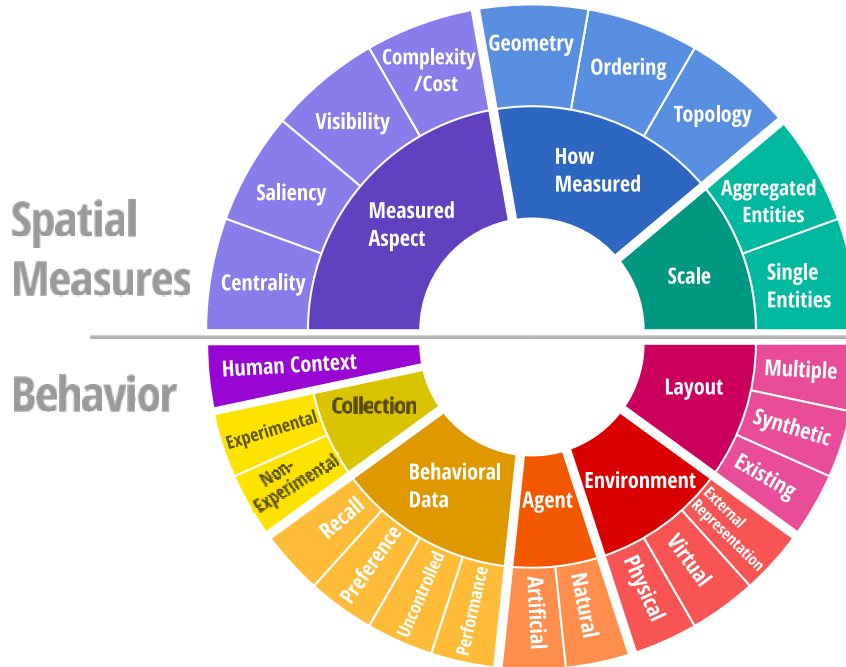
Visual Summary Template

Figure 2: The final visual summary template for spatial measures and human behavior research.

4.1 Category definitions

The categories for the classification are described below, split into two domains: *spatial measures* and *behavior*. In order to ensure coherence in classification among different researchers, it is necessary to define terminology used within the definitions to prescribe how each category is to be applied. For example, the terms *space* and *entity*. Space specifically refers to an area as a whole. This may be on different scales (e.g., a room, a building, a campus, a city, and so on), but it will usually be the encompassing area of a study. Entity,

on the other hand refers to a specific, distinguishable part of a space. This part may be a specific (geographic) object, such as a piece of furniture within a room, a room within a building, a building on a campus, or a district of a city. It can also be generic, referring to location within a space that is not further explicitly specified.

In a case where a paper re-uses data from a previous study, that use is treated identically to if it were original as a practical matter. Research that analyzes data from previous studies or experiments and then also compares it with original data (data collected by the authors specifically for the research conducted in the paper) is classified by taking the pre-existing study or studies into consideration together with the new contribution. An example of this occurs in Turner and Penn [60], in which the authors correlate the results of an artificial agent model to results of a previous study tracking the movement of human agents conducted by Hillier et al. [28]. In the classification, this would result in the paper being categorized as having both artificial and human agent (along with any other applicable category).

4.1.1 Spatial measures

Measured aspect   Beginning at the left of the diagram (Figure 2) and moving clockwise, the first superordinate category within the Spatial Measures domain is *measured aspect*, which represents what the spatial measure within an article means or intends to capture about a space or entity. The subcategories are *centrality*, *saliency*, *visibility*, and *complexity/cost*.

Centrality refers to the importance of an entity based on its spatial relation with other entities. It can be established for structured (e.g., networks) and unstructured spaces. Centrality can be local (based on information of surrounding entities selected by some criteria) or global (based on information of all entities in a space). For example, Baran, Rodríguez, and Khattak [3] compared space syntax measures of local and global centrality to walking behavior in different neighborhoods.

Saliency refers to the distinctiveness or identifiability of an entity, relative to other entities. Nothegger, Winter, and Raubal [46], for example, developed a model that uses various attributes, such as color and facade area, to create a single measure of saliency for buildings.

Visibility represents measures that capture the degree of visibility between one or more entities to or from another entity or set of entities. This is what Benedikt [5] used in creating a series of isovist measures to quantify the nature of visible space around a single point, as well as proposing a series of methods to quantify visibility continuously across space.

Cost/complexity represents how comprehensible the internal structure of a space is (i.e., how easy it is to understand a space) or the mental or physical difficulty of traversing a space (i.e., how easy it is to navigate and/or move through a space to a desired destination). Measures that attempt to measure cost/complexity are often derived from a space's components. O'Neill [49], for example, was interested in the effect of average intersection complexity on wayfinding task performance in relatively small scale spaces (sections of a library). Richter [52] combined different aspects of a road intersection (such as number of branches and segment lengths) to define a measure for that intersection's cognitive complexity.

Note that if authors measure some aspect by combining individual measures that cover other aspects, all aspects are marked in the classification (rather than just the final result of the measure). For example, Nothegger, Winter, and Raubal [46] utilize several environ-

mental aspects, including visibility, to yield a landmark saliency measure. In that case, the paper is considered in the classification to include both saliency and visibility as the measured aspects.

How measured?   *How measured* is the second superordinate category, which captures the actual mathematical structure used to derive values, outside the context of what aspect is being measured. This category is important to classify basic information about methodologies outside the intention of the measure.

The *geometry* category accounts for quantitative (numeric) measuring of geometric properties such as shape, size, angle, and metric distance. Examples include angular change [1, 60], and metric distance [45].

Ordering refers to measures which assess linear/circular order of a finite number of entities, also without measuring metric properties of distance or angle. Richter and Klippel [53], for example, use circular ordering information of an intersection's branches and the position of a landmark object within that order to determine that landmark's location relative to a turn at the intersection (e.g., whether a landmark is located before or after the turn).

The category *topology* represents measures of spatial relationships between entities in terms of connectedness or neighborhood without regard for geometric properties like metric distance, angle, size, or shape. For example, Hölscher, Brösamle, and Vrachliotis [29] identify areas of high centrality, using space syntax measures, which abstract a space into a graph. This graph abstraction allows for the purely topological relationships of the areas within a space to be quantified, such as how connected they are to other areas, without regard for geometric properties like metric distance.

Scale   *Scale*, the last superordinate category within the spatial measures domain, indicates whether a measure operates on an entire space or part of it. Note this is not scale in the sense of map scale, and does not describe the size of the space.

The category of *aggregated entities* captures measures that summarize values from a measure for individual entities into (usually) a single measure for a larger unit. That is, summarizing measures for individual intersections to a single measure for the whole route, or calculating an average measure for a larger area from values for individual locations within that area. An example of this is O'Neill's [49] global ICD, which summarizes the complexity of a space based on aggregating the number of decisions available at all intersections within a space, resulting in a single (average) value for that space.

Single entities, on the other hand, comprises measures which provide values for individual entities of a space and, thus, reveal differences between them. For instance, Baran, Rodríguez, and Khattak [3] use an axial line map and integration to yield a centrality value for each axial line, which defined segments of a pedestrian path network.

4.1.2   Behavior

Behavior is the second domain, representing those categories related to the behavioral portion of the research—as opposed to that which measures aspects of the environment directly.

Human context Beginning from the left and moving counterclockwise, *human context* captures studies which collect additional information on the characteristics of participating human agents, such as familiarity, expertise, sex, or individual differences. Nothegger, et al., [45], for example, had human agents with varying degrees of reported familiarity with the study area identify landmarks at street intersections within the study area. These were then compared to landmarks which were defined as salient by an algorithm constructed for that purpose.

Collection The superordinate category of *collection* captures the two general ways behavioral data is gathered.

The first category is *experimental*, which refers to data recorded by researchers in an original study described in their paper. In papers categorized as such, researchers manipulate variables, such as the environment participants operate in, or the kind of information presented to participants, or the kind of participants, to observe causal relationships. O'Neill [49] once again provides an example of this, in performing an original wayfinding experiment that was conducted with human agents.

In contrast, *non-experimental* captures papers whose methodology is outside the realm of experimental research. In those papers, data is collected without the controlled manipulation of variables, such as with a survey, an observational study, or census data. Koohsari et al. [33] is an example, utilizing mailed surveys that asked human agents to report how often they walked to nearby public open spaces (i.e., parks) and how much time they spent doing so.

Behavioral data The third superordinate category in this domain is *behavioral data*, which stands for the type of data that captures the behavior of agents in some quantifiable way. It contains four categories.

The first is *recall*; evaluations of how well a space and/or entities within that space are remembered by agents. This is measured post-hoc, that is, after some task execution. It may be measured on different levels of spatial knowledge, such as landmark, route, or survey. An example of research fitting into recall comes from Omer and Goldblatt [48], who had human agents mark the location of landmarks on an incomplete map of a space in which they had previously performed wayfinding tasks.

The second category is *preference*, which accounts for research in which agents indicate a preference given a set of choices, as in Weisman [64]. That paper described a task in which human agents judge a series of highly abstracted floor plans in terms of their level of general preference for the plan.

The third category is *uncontrolled*, which represents when research tracks an agent's decisions without a specific goal imposed upon the agents by the researchers. The researchers do not measure performance (of any kind), but simply observe behavior, such as how many agents pass by or enter a specific location, such as in Chang [12], who observed pedestrian movement by creating "gates" at particular points in a physical space and measured how many human agents passed through them. In addition, human agents passing through the gates were randomly selected and their movement tracked through the space.

The last category within behavioral data is *performance*. In contrast to recall, this measures the observable performance *during* task execution, such as time to completion, route optimization, or number of turning errors. O'Neill [49] contains an example of gathering performance-type data. Human agents were used in a wayfinding experiment, and their

ability to find a predetermined destination was measured in three ways: 1) time elapsed, 2) number of backtracks on route, and 3) number of wrong turns.

Agent   The superordinate category *Agent* accounts for research using different beings whose behavior or actions are observed (and compared to particular spatial measures).

Natural captures research which uses human agents, people used as test participants, as in O'Neill [49], which had students perform wayfinding tasks.

On the other hand, *artificial*, represents research which uses agents whose actions are intended to approximate human behavior, as in Turner and Penn [63], who used an agent-based model in which the agents used visibility information to make movement decisions.

Environment   *Environment* is the next superordinate category and represents the different ways environment/spaces are experienced in the study by agents.

First, research can use *physical* spaces, an actual physical space as it exists in the real world, as in O'Neill [49], who had agents perform tasks in a university library building.

Second, there are *virtual* environments, which are computer-generated spaces with or without the rules of physical reality, such as a digital three-dimensional model of a campus. One example of this is Meilinger, Franz, and Bülthoff [42], who had human agents experience a three-dimensional simulation of a town through an immersive 220° semi-cylindrical screen.

The last environment category is *external representation*. This represents studies that utilize an (often static) representation of a space, such as a floorplan, map, or series of photographs, in addition to (or instead of) having agents interact in or move within a real or simulated environment. Asami et al. [1] had designated experts select local centers in Istanbul using a map of the city. O'Neill [49] had both "physical" and "external representation" environments. Human agents first experienced the environment through the use of a series of photographs in order to familiarize them with the space before encountering it physically.

Layout   The last superordinate category within behavior is *layout*, which captures the type of spatial structure or arrangement of the environment used within a paper.

The first is *existing*, which refers to a spatial layout found in the real world that could be encountered in life, such as a street grid of a real place. For example, Hölscher, Brösamle, and Vrachliotis [29] had human agents perform tasks within a pre-existing building (a conference center), while Meilinger, Franz, and Bülthoff [42] had human agents experience a *virtual* model of the town of Tübingen, Germany.

The second category is *synthetic*, which stands for when a paper uses a layout that is designed specifically to test behavioral responses to environmental aspects, such as a maze or idealized rectilinear street network. Dalton [15] did this by creating a virtual, idealized urban form consisting of streets of all the same length, but a variety of intersection types, which human agents then experienced through a computer simulation.

Rather than standing for a different type of layout, the category of *multiple* captures studies that compare multiple layouts, that is, multiple different spaces (or multiple distinct areas within a space). In that vein, O'Neill [49] compared three different areas within a library which varied in measured complexity, in order to validate the measure against human behavior.

4.2 Considered categories

Over the course of affinity diagramming many categories were created, deleted, and revised. Importantly, there were categories we expected to find, and ways we anticipated to structure the taxonomy that did not hold up under a close reading of the corpus. Some absent categories seem to indicate there are under-researched areas that can be identified with this method. If they begin to materialize in the literature, the taxonomy and visual summary diagram can be updated accordingly (a possibility further discussed in Section 5).

One notable example of a conceptually sound category that was not present in the literature was three-dimensional spatial measures. Three dimensional measures seem to be a neglected aspect of spatial measures research, especially in the face of the increasing commonality of three-dimensional geospatial data, particularly for urban environments [34]. While many of the papers attempted to capture three-dimensional information, this generally entailed abstracting the environment so that a two-dimensional measure could be used. An example of this can be seen in Hölscher, Brösamle, and Vrachliotis [29], who analyzed multiple floors of a building together by creating "dummy" connections to simulate them all being on the same level.

Another example was *dynamic layouts*. Currently all encountered research seems to assume that effectively an environment is static, or at least does not meaningfully change during a task such as navigation. Conceptually, it is not difficult to imagine an environment where the navigable space changes quickly, such as a maze with moving walls. Obviously, this would be easier to implement experimentally in a virtual environment, but there are real-life situations where the environment is dynamic, such as a street network prone to accidents, or a burning building.

Some of the discarded categories had roots in common sense, but upon closer examination were too difficult to define satisfactorily. We make the assumption that if we cannot effectively create a logically consistent definition, then a user will have an even harder time parsing or applying it. Several proto-categories cut for this reason fell under a superordinate category titled *abstraction* (or *entity type*). This would have referred to the base data type of the representation, or the type of abstraction from the real world. Lynch [39], for example, distinguishes paths, edges, districts, nodes, and landmarks. Golledge [22] provides geometric components of spatial knowledge as points, lines, areas, and surfaces. While common conceptually, actually distinguishing the different types within the literature is nontrivial. Several common measures operate partly by transforming between such abstractions, such as axial line mapping in space syntax, which uses a linear representation to then define a network on which graph measures are calculated [27]. In a similar vein, isovist fields [4], account for various polygonal metrics for visibility at every point in a space.

We were cognizant of the difficulty in creating objective criteria for assigning categories given these complexities, and then informing users of these criteria. Accordingly, it was decided to instead focus on the intent of the measures and only provide high level information in regards to methodology. This is represented in the *measured aspect* and *how measured* superordinate categories, respectively. Another revision was to change *global* and *local* in *scale* to *aggregated entities* and *single entities*. The basic problem was that distinguishing between purely local and global depends on the reference frame, which would then also need to be indicated in the category structure and visual summary. This was in fact also considered, but seemed to entail strict but rather arbitrary delineations of particular

scales. For example, when does one move from a neighborhood to a district, or from a city to region?

4.3   Validation of categories

In order to verify the applicability of the proposed categorization, we undertook three different literature classification tasks for which inter-rater agreement was quantified via Cohen's Kappa [13]. Cohen's Kappa accounts for agreement that could be expected by chance alone. Two rounds of classification were completed among three of the authors, with a third round using one author and one outside "naïve" rater only somewhat familiar with the research area (a cognitive neuroscientist).

Kappa Statistic	< 0.00	0.00–0.20	0.21–0.40	0.41–0.60	0.61–0.80	0.81–1.00
Strength of Agreement	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect

Table 1: Guidelines for interpreting Cohen's Kappa values [25].

Author-only classifications directly informed revisions to the categories. Ten papers were independently classified in the first round, using a set of preliminary definitions created in the affinity diagramming method described in Section 3.1. The authors' results were compared pairwise, with the following Cohen's Kappa values for each of the three pairs: 0.66, 0.65, and 0.60. These values fall between "Moderate" and "Substantial" according the guidelines for interpreting Cohen's Kappa provided by Landis and Koch [35] (see Table 1). This indicated the classification was not mature, and repeated disagreements, such as consistent differences for particular categories, were analyzed and discussed, and the classification revised. Some of the details of this discussion and its results are examined in Section 4.2. Once the revision of the categories was complete, a second round using three papers was completed. This was considerably more successful, with two pairs recording identical Cohen's Kappa of 0.80 (extreme high end of "substantial") and one pair returning a result of 0.94 ("almost perfect") [35]. After a discussion, which focused on understanding the remaining disagreements, and small follow-up revisions to definitions including examples for each category, the naïve rater evaluation began.

In the naïve rater task one author and a naïve rater categorized five papers: [30, 31, 37, 38, 65]. The naïve rater was a cognitive neuroscientist, previously unfamiliar with the classification, but somewhat familiar with basic research in the area. Both raters were given the set of full definitions (see Appendix B: Category Definitions for the document used), which included instructions covering particular scenarios that might be encountered, such as authors using existing data to test a new method. Care was taken to select papers that had not been used as examples within the definitions document. Additionally, an example classification of O'Neill's paper [49] was provided, complete with a visual summary, seen in Figure 3. Once again, agreement was measured with Cohen's Kappa, which resulted in a high value of 0.83 (the breakdown of the classification is shown in Table 2). While only using one naïve rater may not be ideal, we consider this to be a highly satisfactory result that illustrates the categories identified are not idiosyncratic, but are generalizable ways of understanding this research area that are also usable by non-experts.

Paper:		Wiener [22]	Li [27]	Holscher [28]	Jiang [50]	Lu [23]
Measured Aspect	Centrality	Blue	Orange	Blue	Orange	Blue
	Saliency	Blue	Blue	Blue	Blue	Blue
	Visibility	Green	Green	Orange	Blue	Green
	Cost/Complexity	Blue	Blue	Blue	Orange	Blue
How Measured	Topology	Green	Green	Green	Green	Green
	Ordering	Blue	Blue	Blue	Blue	Blue
	Geometry	Green	Blue	Green	Blue	Blue
Scale	Aggregated Entities	Orange	Green	Green	Green	Blue
	Single Entities	Orange	Green	Blue	Green	Green
Agent	Natural	Green	Green	Green	Green	Green
	Artificial	Blue	Blue	Blue	Orange	Blue
Behavioral Data	Recall	Blue	Orange	Blue	Blue	Blue
	Preference	Green	Blue	Green	Blue	Blue
	Uncontrolled	Blue	Blue	Blue	Green	Green
	Performance	Green	Blue	Blue	Green	Green
Collection	Experimental	Green	Green	Blue	Green	Green
	Non-Experimental	Blue	Blue	Blue	Green	Green
Environment	Physical	Blue	Green	Blue	Green	Green
	Virtual	Green	Blue	Orange	Blue	Blue
	External Representation	Blue	Blue	Blue	Blue	Blue
Layout	Existing	Blue	Green	Green	Green	Green
	Synthetic	Green	Green	Blue	Blue	Blue
	Multiple	Green	Green	Green	Green	Blue
Human Context		Blue	Green	Green	Blue	Orange

Table 2: Results of the naïve rater task. Green shows agreement on assigning a paper to a category, blue shows agreement on not assigning a paper to a category, and orange shows a disagreement on whether or not to assign a paper to a category.

4.4 Example visual summary

We return to the example of O'Neill [49] to demonstrate how our taxonomy functions in practice, with the results illustrated in Figure 3. O'Neill measured the performance and memory of students when wayfinding in a library, compared with complexity of the environment as defined by a simple network complexity measure called ICD. Note that the appendix shows visual summaries for 16 other papers that have been classified.
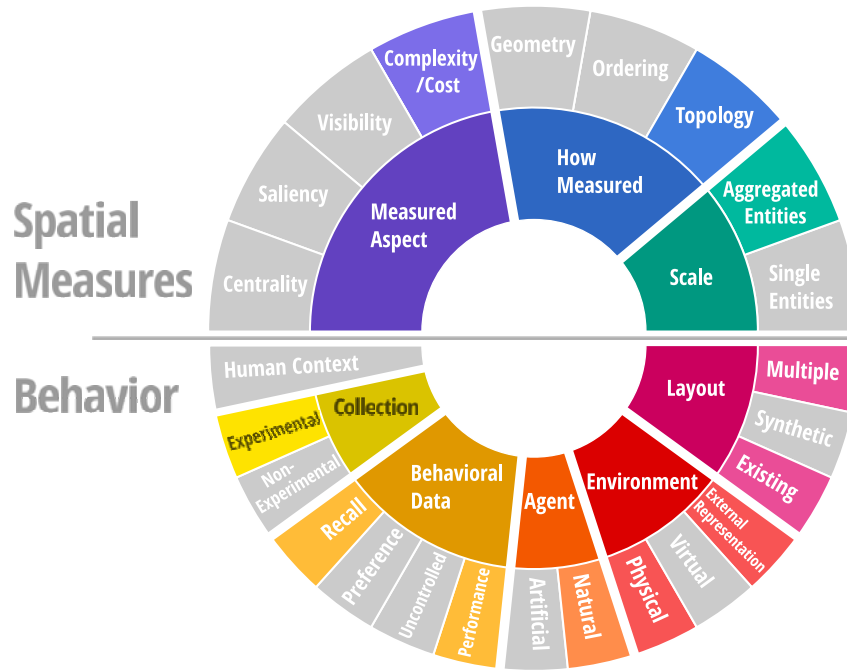
O'Neill 1991

Figure 3: An example visual summary created for O'Neill [49].

Beginning from the top left of the visual summary (Figure 3), this paper has *complexity/cost* as a *measured aspect*, as the ICD measure O'Neill uses is intended to summarize the complexity of the environment. ICD does this by simple topology, summarizing the degree of nodes (intersections; termed *choice points*) in the floorplan of the environments. Thus, the *topology* and *how measured* categories are highlighted. Since the measure aggregates the values of all the choice points in an environment rather than singling out any individual value, *scale* and *aggregated entities* are highlighted. For *layout*, *multiple* is flagged because three different environments within the library (with different complexity) were evaluated. *Existing* is flagged because these environments already existed and were not designed for the task by the researcher. For *environment*, the library settings were physical spaces (as opposed to virtual), thus *physical* is highlighted. *External representation* is also flagged because O'Neill used a series of photographs to give a "guided tour" of the environment to participants as a pre-training task before the wayfinding component of the experiment.

The participants were graduate and undergraduate students, so *natural* is marked under *agent*. Within *behavioral data*, both *performance* and *recall* are marked. The former because the study recorded performance in the wayfinding task in the form of time to task completion, number of backtracks, and number of incorrect turns compared to an optimal route. The latter is flagged because O'Neill also had participants draw a sketch map of the envi-

ronment they navigated, and evaluated that map for completeness and accuracy. O'Neill explicitly conducted an experiment (rather than a survey or other method), so *collection* and *experimental* are flagged. Since O'Neill did not compare any aspect of the participants (such as age, gender, or experience) other than the results of their experimental task, *human context* remains un-flagged.

4.5 Web-based classification and analysis tool

In order to simultaneously bring our results to a wider audience of scholars and to build our knowledge base, we are in the final stages of implementing a web interface for both exploring and creating visual summaries.[1] Once a user is registered, they can create visual summaries for new papers and add their bibliographic information via BibTex [18], which is then added to the corpus database. The intent is that this approach will function as a form of crowdsourcing, increasing the size (and therefore utility) of our corpus, and opening up opportunities for future analysis.

Users can browse the existing database of visual summaries, and select papers by category, so that only papers with the selected categories are displayed. An example use-case of this feature is given in Section 4.6. In addition to manual visual comparison, the website also has an interactive overview visual summary. Intended to give users a sense of larger trends, this displays how often categories are present within the entire corpus database, or a selection of that database. A screenshot of the current status is shown in Figure 4. Darker segments indicate more popular categories.

Currently, users can select to show only papers within a date range (by year of publication), which can give an indication of trends by simply changing the selected range. More selection features are planned, such as being able to select by author name.

4.6 Example use cases

One major strength of using a diagram to represent categories is that it makes comparing literature straightforward. An example use case would be a researcher who is familiar with a specific body of research, but looking to explore whether a specific idea has been pursued outside that body. Using our web interface, they could select their topics of interest. For example, they could be interested in the visibility of landmarks in virtual reality, and select the appropriate subordinate categories (visibility and virtual). In the results, they could look for unfamiliar names or unusual combinations of categories (such as the use of non-experimental data collection).

Let us now take the perspective of another user, a first-semester graduate student who is interested in work that examines navigational difficulty based on environment complexity for the purposes of planning an experiment to base their thesis around. Using our website, they query for literature that has *complexity/cost* measures and *experimental* data collection. The visual summaries in Figure 5 are shown: O'Neill [49], Turner and Penn [63], Dalton [14], Omer and Goldblatt [48], and two papers by Hölscher with different co-authors [29, 30].

Several patterns can be noticed immediately within the selected categorizations. First, *complexity/cost* is often investigated jointly with *visibility*, but never with *centrality*. Ordering

---

[1]For the current state of implementation, see http://chorophronesis-magicat.rhcloud.com/visualsummary/home/
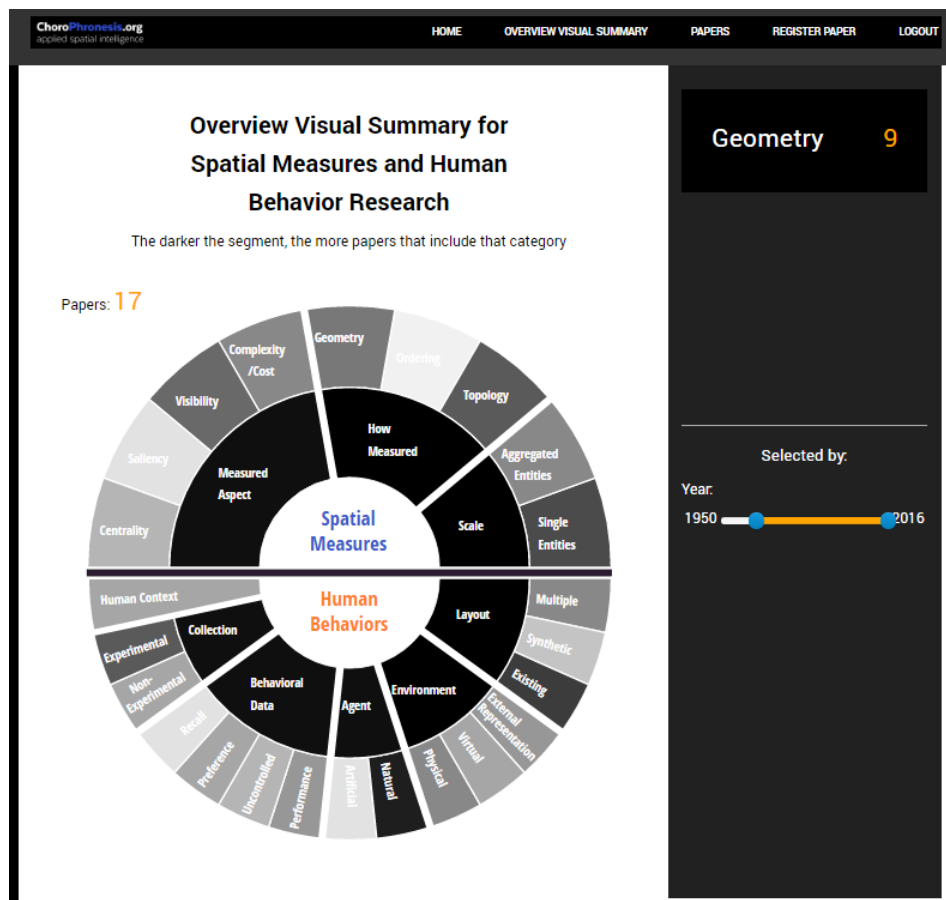
Figure 4: Screenshot of web interface. The overview visual summary for all papers, as shown in our web interface (in progress). The darker the segment, the more papers that contain that category. Users can mouse over segments to get an exact count of papers that include it (displayed top right). Users can also use an interactive timeline to restrict the display by publication year. Other features are planned, such as query by author name.

never appears, perhaps indicating that using it as an approach could be novel in this context. Most research uses *aggregated entities*, but not all. All collection is *experimental* (since this was one of the query terms), but one paper supplements this with *non-experimental* data collection. The actual type of behavioral information collected varies considerably, perhaps indicating to our graduate student that they need to consider the type of data they would like to collect to refine their search (such as focusing on *performance*). *Natural* agents (people) might be expected, but perhaps the only paper to use *artificial* agents (Turner and Penn's) might provide some insight as to how they can be employed. Similar to behavioral data used, the type of environment varies, indicating another area that requires some thought for further specification. Half of the papers examine both *multiple* and *existing* layouts, which may indicate a logical direction for evaluation (as it appears to be a popular
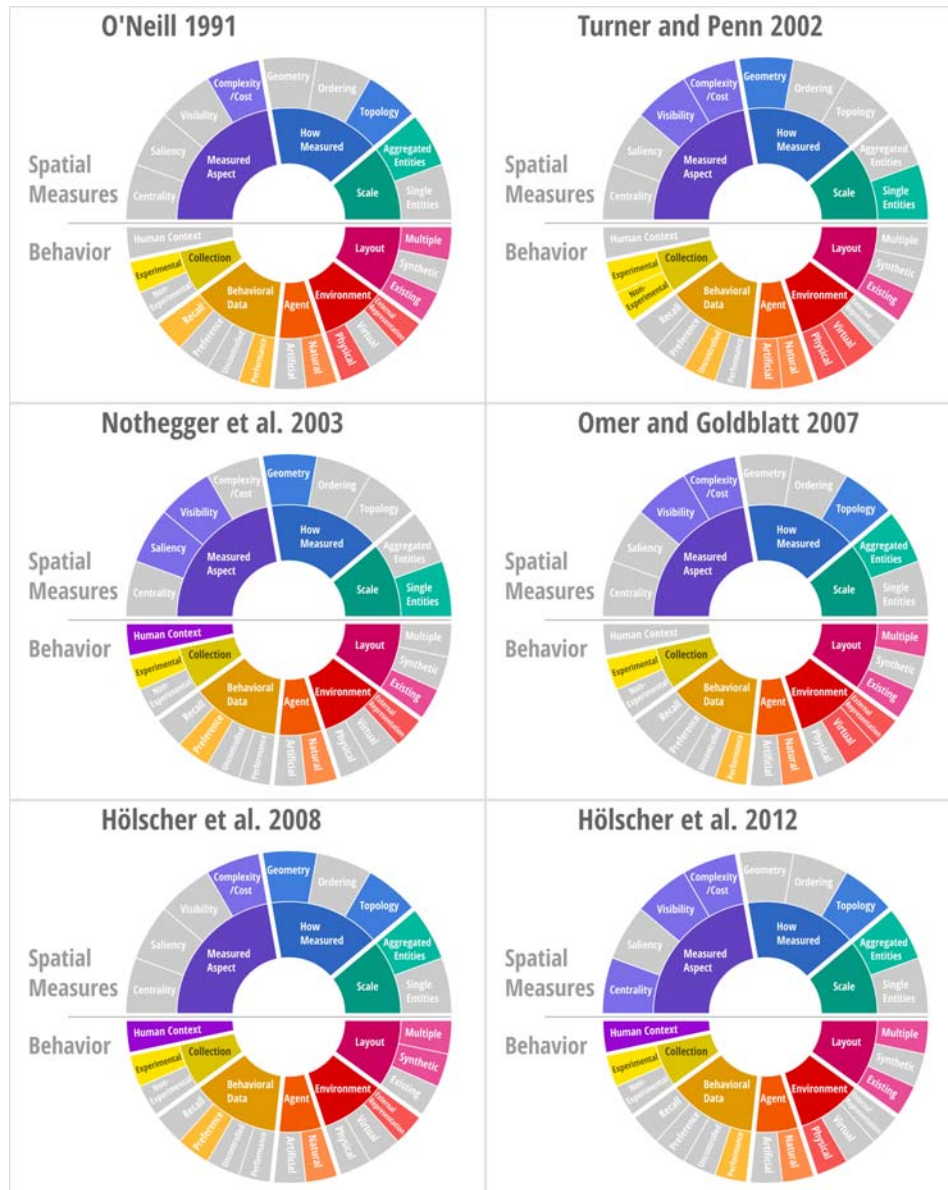
Figure 5: An illustration of how multiple visual summaries can be used to compare papers quickly. All papers share the categories of Complexity/Cost and Experimental (collection).

theme). However, the lack of papers that have attempted to use or compare *existing* and *synthetic* layouts points to what could be the basis for an interesting experiment. Outside of the categories themselves, our student could note that Hölscher has at least two papers and both are relatively recent, so his work might be especially relevant.

While this example may be a somewhat idealized scenario, we believe it nicely illustrates the possibilities of our visual summary output for identifying themes and potential avenues for future research.

4.7   Why not automated text analysis?

The modified affinity diagramming method used by Mason et al., and modified here, is used in lieu of automated text analysis methods. This is primarily because we feel they would not function well for our particular purpose: creating a robust taxonomy for a highly interdisciplinary domain whose core literature set is not intractably large. Automated text analysis methods still require humans to provide domain knowledge [21, 23], whether they are used to identify topics within literature or assign topics (categories) to literature. In unsupervised classification a person is still needed to make sense of the output of discovered patterns, such as nominal topics [43]. Our analysis, i.e., the creation of categories, is most closely analogous to *topic modeling*, the various approaches aimed at automatically extracting "topics" from text corpora. For overviews on topic modeling, see, for example, Mohr and Bogdanov [43] and Brett [10].

Similarly, automated assignment of categories of literature (a classification problem) requires a person to define the categories to apply, and often to classify papers in order to create a training set for the model [23]. Given that we are faced with a corpus that is both comparatively small and rather heterogeneous, the effort needed to fine-tune automated methods or training sets for them to produce sensible categorizations seems unnecessary and wasted, particularly given that they would still be unlikely to capture all intricacies and seeming contradictions in term use, for example. Theoretically the affinity diagramming method does not even require the use of a computer, which is advantageous where resources are limited or there is a lack of technical expertise. Still, the use of automated methods for visual summaries may prove to be especially useful in some scenarios, an idea we discuss in Section 5.

The advantages of the combination of affinity diagramming and visual summary can further be seen when compared with a word cloud, a basic form of text analysis. A word cloud displays common words from a text source, with the larger font used corresponding to more common words (barring common words in English syntax). An example word cloud created using www.wordle.net [19] is shown in Figure 6. The text used was the abstracts from a selection of our corpus (the same selection is used in Appendix A: Additional Visual Summaries).

A word cloud is a simplistic form of text analysis, but it clearly illustrates a fundamental challenge: the difference between words (as a group of characters) and their actual meaning. For example, "graph" and "network" can be synonymous, or have completely different meanings, depending on the context (i.e., a statistical graph). This specific example is also discussed by Borrett et al. [3], who in the context of analyzing network ecology literature had to adjust for multiple keywords and spurious phrases that included related terms but were not directly of interest, such as "transportation network."

This disconnect between literal text and meaning is an issue particularly for interdisciplinary work such as ours, where multiple conventions are in play. The variety of terminology used within spatial cognition literature alone can be quite impressive: Golledge [8] lists no less than twenty-one alternative terms for the concept of an internal representation present in literature on spatial cognition. The value of a category structure created manu-

Figure 6: A word cloud made using abstracts of literature within our corpus.

ally, especially in an interdisciplinary context, is that it inherently allows content and ideas to be disconnected from the specific language used to describe them. This aspect cannot currently be achieved with automated text analysis approaches of any kind.

4.8   Visual summary design

The design of the visual summary can be seen in Figure 2. According to Mackinlay [40], the two most effective visual variables [6] for "perceptual tasks" using nominal data are position and color hue. We focus on these two elements to represent the categorical meaning of the elements in the diagram, while attempting to control for other visual variables, such as size, and maintaining an aesthetic balance.

First, the diagram is split into two halves: the top half containing spatial measures categories and the bottom half containing human behavior categories. The summary is hierarchical from the inside out, so that the outside categories (represented with "slices") are subordinate to the inside categories. While the number of categories contained within each half is different, we felt it was important to keep the halves the same overall size to avoid the user assuming differing importance. Related category segments are grouped together, but not ordered. The inner ring of segments contains the superordinate categories, while the outer ring contains subordinate categories. The sizes of the segments are determined by the subordinate categories, so that all subordinate categories are the same size within each half. Superordinate categories are then sized to match the number of subordinate categories they contain. This leads to some superordinate categories appearing bigger than others. This may imply differing importance to some, but on the other hand ensures that subordinate categories—which are more salient—keep the same size, and text in the diagram remains readable while the overall diagram still has a reasonable size.

Color hue is the other major visual variable used, with "warm" (more yellow or red) colors for human behavior visual elements, and "cool" (blue and green) colors for visual elements relating to spatial measures. Colors within those ranges are then assigned to cate-

gory groups, with all the subordinate categories having the same color (to avoid indicating importance), and the superordinate category segment having a slightly darker tone. The circular shape was chosen to allow for segments to be added and removed in case of future revisions, without a need for a complete redesign. For example, if *dynamic* environments as discussed in Section 4.2 were to be covered in future literature, they may be added to the "Environment" slice.

The visualization was refined and designed as the categories were refined. The emphasis of the design is on clarity of information. Having a working visualization also helped to find the right level of granularity of representation for the intended purpose. Too general and the diagram fails to provide useful information; too detailed and users will have difficulty finding what they are looking for. Two levels of hierarchy were selected over more detailed alternatives, since we believe this is a reasonable balance between visual search difficulty and level of detail.

5   Future work

The inter-rater agreement tasks speak to the validity of our taxonomy, and we believe the visual summary provides a valuable way to access and understand that categorization. However, there are opportunities for future research, both in terms of extending and applying our existing visual summary and taxonomy, and improving the method.

As discussed in Section 4.5, we are finalizing the implementation of a web interface for our visual summary of spatial measures and human behavior research. The site will allow scholars to view existing visual summaries, get an overview of frequency of categories in the database, and create their own visual summary for a given paper, which automatically adds that paper to the database. While currently the analytical features are still limited, we are planning to implement more advanced tools. There are many classic bibliometric analyses that would be interesting to combine with the categorical information of our taxonomy. For instance, which journals or authors are associated with which categories, and how those associations change over time.

The web interface offers more opportunities for further evaluations of the taxonomy and the visual summary. For example, crowdsourcing-style evaluations of the taxonomy could be performed by asking multiple users to classify the same papers and then comparing their mutual agreement across papers and categories. In theory, an automated system for creating visual summaries makes creating different representations of the same data relatively easy, making an evaluation of different visualization methods possible.

We have made a considerable effort to identify intuitive categories that currently exist in spatial measures and human behavior literature, however we fully expect new categories to be needed to account for future research directions. Any changes will require revision of the taxonomy, visual summary diagram, and the web interface. In Section 4.2 we identified several categories we expected to find but did not, pointing to potential research directions. No doubt continued use of the website will bring new possibilities to light in time. Fortunately, the visual summary diagram was designed with the possibility of adding categories through additional segments, as we noted in Section 4.8. A deeper challenge will be re-classifying our existing corpus to account for the new categories. However, it seems that many potential categories will be completely new in more recent work, so that past papers will simply not have that category assigned. Existing visual summaries can then be

updated automatically within the web interface. In the unlikely event that new categories are found which require a complete review of the existing corpus, a manual review would be necessary, which would be time consuming but not impractical.

While we have discussed the advantages of using the modified affinity diagramming method over automated text mining, we believe there might be potential in exploring machine learning methods to support the categorization process. For example, cluster analysis could be used to help find previously unrecognized subfields of research, or jump start the categorization process, especially for broader or more complex fields. Despite the potential of extending the method, we believe the current iteration provides a helpful, easy to comprehend way to both create an understanding of a field of literature (through the creation of categories) and communicate that understanding in an effective way that enhances readers' understanding of single papers and entire fields.

Acknowledgments

References

[1] ASAMI, Y., KUBAT, A. S., KITAGAWA, K., AND IIDA, S.-I. Introducing the third dimension on space syntax: Application on the historical Istanbul. In *Proc. 6th International Space Syntax Symposium* (Istanbul, Turkey, 2007), A. S. Kubat, Ed., ITU Faculty of Architecture, pp. 48.1–48.18.

[2] BAFNA, S. Space syntax: A brief introduction to its logic and analytical techniques. *Environment and Behavior 35*, 1 (2003), 17–29. doi:10.1177/0013916502238863.

[3] BARAN, P. K., RODRÍGUEZ, D. A., AND KHATTAK, A. J. Space syntax and walking in a new urbanist and suburban neighbourhoods. *Journal of Urban Design 13*, 1 (2008), 5–28. doi:10.1080/13574800701803498.

[4] BATTY, M. Exploring isovist fields: Space and shape in architectural and urban morphology. *Environment and Planning B: Planning and Design 28*, 1 (2001), 123–150. doi:10.1068/b2725.

[5] BENEDIKT, M. L. To take hold of space: Isovists and isovist fields. *Environment and Planning B: Planning and Design 6*, 1 (1979), 47–65. doi:10.1068/b060047.

[6] BERTIN, J. *Semiology of Graphics: Diagrams, Networks, Maps*, 1st ed. ed. ESRI Press and Distributed by Ingram Publisher Services, Redlands, Calif., 2010.

[7] BÖRNER, K., CHEN, C., AND BOYACK, K. W. Visualizing knowledge domains. *Annual Review of Information Science and Technology 37*, 1 (2003), 179–255. doi:10.1002/aris.1440370106.

[8] BORRETT, S. R., MOODY, J., AND EDELMANN, A. The rise of network ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling 293* (2014), 111–127. doi:10.1016/j.ecolmodel.2014.02.019.

[9] BOYACK, K. W., KLAVANS, R., AND BÖRNER, K. Mapping the backbone of science. *Scientometrics 64*, 3 (2005), 351–374. doi:10.1007/s11192-005-0255-6.

[10] BRETT, M. R. Topic modeling: A basic introduction. *Journal of digital humanities 2*, 1 (2012), 12.

[11] CARLSON, L. A., HÖLSCHER, C., SHIPLEY, T. F., AND DALTON, R. C. Getting lost in buildings. *Current Directions in Psychological Science 19*, 5 (2010), 284–289. doi:10.1177/0963721410383243.

[12] CHANG, D. Spatial choice and preference in multilevel movement networks. *Environment and Behavior 34*, 5 (2002), 582–615. doi:10.1177/0013916502034005002.

[13] COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20*, 1 (1960), 37–46. doi:10.1177/001316446002000104.

[14] DALTON, R. C. The secret is to follow your nose: Route path selection and angularity. *Environment & Behavior 35*, 1 (2003), 107–131. doi:10.1177/0013916502238867.

[15] DALTON, R. C., HÖLSCHER, C., AND TURNER, A. Understanding space: The nascent synthesis of cognition and the syntax of spatial morphologies. *Environment and Planning B: Planning and Design 39*, 1 (2012), 7–11. doi:10.1068/b3901ge.

[16] DARA-ABRAMS, D. Cognitive surveying: A framework for mobile data collection, analysis, and visualization of spatial knowledge and navigation practices. In *Spatial cognition VI*, C. Freksa, Ed., vol. 5248 of *Lecture notes in computer science, 0302-9743*. Springer, Berlin, 2008, pp. 138–153. 10.1007/978-3-540-87601-4doi:\textunderscore 12.

[17] EVANS, G. W. Environmental cognition. *Psychological Bulletin 88*, 2 (1980), 259–287. doi:10.1037/0033-2909.88.2.259.

[18] FEDER, A. Bibtex. http://www.bibtex.org/, 2016. Last accessed 10 December 2016.

[19] FEINBERG, J. Wordle - create. http://www.wordle.net/create, 2016. Last acccessed May 4 2016.

[20] FRANZ, G., AND WIENER, J. M. From space syntax to space semantics: A behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment and Planning B: Planning and Design 35*, 4 (2008), 574–592. doi:10.1068/b33050.

[21] FUCHS, G., STANGE, H., SAMIEI, A., ANDRIENKO, G., AND ANDRIENKO, N. A semi-supervised method for topic extraction from micro postings. *it - Information Technology 57*, 1 (2015). doi:10.1515/itit-2014-1078.

[22] GOLLEDGE, R. G. *Wayfinding Behavior: Cognitive Mapping and other Spatial Processes*. Johns Hopkins University Press, Baltimore and London, 1999.

[23] GRIMMER, J., AND STEWART, B. M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis 21*, 3 (2013), 267–297. doi:10.1093/pan/mps028.

[24] HALLOWELL, G. D., AND BARAN, P. K. Suburban change: A time series approach to measuring form and spatial configuration. *The Journal of Space Syntax 4*, 1 (2013), 74–91.

[25] HARDY, A. C. Landscape and human perception. *Methods of Landscape Analysis October* (1967), 3–8.

[26] HILLIER, B. *Space is the Machine: A configurational theory of architecture*. Cambridge University Press, Cambridge, 1996.

[27] HILLIER, B., AND HANSON, J. *The Social Logic of Space*. Cambridge University Press, Cambridge, 1984.

[28] HILLIER, B., MAJOR, M. D., DESYLLAS, J., KARIMI, K., CAMPOS, B., AND STONOR, T. Tate Gallery, Millbank: A study of the existing layout and new masterplan proposal. Tech. rep., Bartlett School of Graduate Studies, University College London, London, UK, 1996.

[29] HÖLSCHER, C., BRÖSAMLE, M., AND VRACHLIOTIS, G. Challenges in multilevel wayfinding: A case study with the space syntax technique. *Environment and Planning B: Planning and Design 39*, 1 (2012), 63–82. doi:10.1068/b34050t.

[30] HÖLSCHER, C., AND DALTON, R. C. Comprehension of layout complexity: Effects of architectural expertise and mode of presentation. In *Design Computing and Cognition '08*, J. S. Gero and A. K. Goel, Eds. Springer, Dordrecht and London, 2008, pp. 159–178. 10.1007/978-1-4020-8728-8doi:\textunderscore 9.

[31] JIANG, B. Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science 23*, 7 (2009), 823–837. doi:10.1080/13658810802022822.

[32] KOHONEN, T. Essentials of the self-organizing map. *Neural networks 37* (2013), 52–65. doi:10.1016/j.neunet.2012.09.018.

[33] KOOHSARI, M. J., KACZYNSKI, A. T., GILES-CORTI, B., AND KARAKIEWICZ, J. A. Effects of access to public open spaces on walking: Is proximity enough? *Landscape and Urban Planning 117* (2013), 92–99. doi:10.1016/j.landurbplan.2013.04.020.

[34] LAFARGE, F., AND MALLET, C. Creating large-scale city models from 3d-point clouds: A robust approach with hybrid representation. *International Journal of Computer Vision 99*, 1 (2012), 69–85. doi:10.1007/s11263-012-0517-8.

[35] LANDIS, J. R., AND KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics 33*, 1 (1977), 159. doi:10.2307/2529310.

[36] LARKIN, J., AND SIMON, H. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science 11*, 1 (1987), 65–100. doi:10.1016/S0364-0213(87)80026-5.

[37] LI, R., AND KLIPPEL, A. Wayfinding behaviors in complex buildings: The impact of environmental legibility and familiarity. *Environment and Behavior 48*, 3 (2016), 482–510. doi:10.1177/0013916514550243.

[38] LU, Y., AND ZIMRING, C. Can intensive care staff see their patients? An improved visibility analysis methodology. *Environment and Behavior 44*, 6 (2012), 861–876. doi:10.1177/0013916511405314.

[39] LYNCH, K. *The Image of the City*. MIT Press, Cambridge, MA, 1960 (1975).

[40] MACKINLAY, J. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics 5*, 2 (1986), 110–141. doi:10.1145/22949.22950.

[41] MASON, J. S., RETCHLESS, D., AND KLIPPEL, A. Domains of uncertainty visualization research: A visual summary approach. *Cartography and Geographic Information Science* (2016), 1–14. doi:10.1080/15230406.2016.1154804.

[42] MEILINGER, T., FRANZ, G., AND BÜLTHOFF, H. H. From isovists via mental representations to behaviour: First steps toward closing the causal chain. *Environment and Planning B: Planning and Design 39*, 1 (2012), 48–62. doi:10.1068/b34048t.

[43] MOHR, J. W., AND BOGDANOV, P. Introduction—topic models: What they are and why they matter. *Poetics 41*, 6 (2013), 545–569. doi:10.1016/j.poetic.2013.10.001.

[44] MONTELLO, D. R. The contribution of space syntax to a comprehensive theory of environmental psychology. In *Proc. 6th International Space Syntax Symposium* (Istanbul, Turkey, 2007), A. S. Kubat, Ed., ITU Faculty of Architecture, pp. iv.1–iv.12.

[45] NAGAR, A., AND TAWFIK, H. A multi-criteria based approach to prototyping urban road networks. *Issues in Informing Science and Information Technology Education 4* (2007), 749.

[46] NOTHEGGER, C., WINTER, S., AND RAUBAL, M. Selection of salient features for route directions. *Spatial Cognition and Computation 4*, 2 (2004), 113–136. 10.1207/s15427633scc0402doi:\textunderscore1.

[47] NUBANI, L., AND WINEMAN, J. The role of space syntax in identifying the relationship between space and crime. In *Proc. International Space Syntax Symposium* (Amsterdam, Netherlands, 2005), A. van Nes, Ed., pp. 413–422.

[48] OMER, I., AND GOLDBLATT, R. The implications of inter-visibility between landmarks on wayfinding performance: An investigation using a virtual urban environment. *Computers, Environment and Urban Systems 31*, 5 (2007), 520–534. doi:10.1016/j.compenvurbsys.2007.08.004.

[49] O'NEILL, M. J. Evaluation of a conceptual model of architectural legibility. *Environment and Behavior 23*, 3 (1991), 259–284. doi:10.1177/0013916591233001.

[50] PAUL, A. A comparative assessment of edge-effect with syntax integration generated in axial and unit-segment approaches to modelling vehicular movement networks. *International Journal of Urban Sciences 18*, 3 (2014), 340–354. doi:10.1080/12265934.2014.908131.

[51] PEPONIS, J., ZIMRING, C., AND CHOI, Y. K. Finding the building in wayfinding. *Environment and Behavior 22*, 5 (1990), 555–590. doi:10.1177/0013916590225001.

[52] RICHTER, K.-F., AND DUCKHAM, M. Simplest instructions: Finding easy-to-describe routes for navigation. In *Geographic Information Science*, T. J. Cova, Ed., vol. 5266 of *Lecture notes in computer science, 0302-9743*. Springer, Berlin, 2008, pp. 274–289. 10.1007/978-3-540-87473-7doi:\textunderscore 18.

[53] RICHTER, K.-F., AND KLIPPEL, A. Before or after: Prepositions in spatially constrained systems. In *Spatial cognition V*, T. Barkowsky, Ed., vol. 4387 of *Lecture notes in computer science. Lecture notes in artificial intelligence, 0302-9743*. Springer, Berlin, 2007, pp. 453–469. 10.1007/978-3-540-75666-8doi:\textunderscore 26.

[54] SKEELS, M., LEE, B., SMITH, G., AND ROBERTSON, G. G. Revealing uncertainty for information visualization. *Information Visualization 9*, 1 (2009), 70–81. doi:10.1057/ivs.2009.1.

[55] SKUPIN, A., AND AGARWAL, P. Introduction: What is a self-organizing map? In *Self-organising maps*, P. Agarwal and A. Skupin, Eds. Wiley, Chichester, England and Hoboken, NJ, 2008, pp. 1–20. doi:10.1002/9780470021699.ch1.

[56] SKUPIN, A., BIBERSTINE, J. R., AND BORNER, K. Visualizing the topical structure of the medical sciences: a self-organizing map approach. *PloS one 8*, 3 (2013), e58779. doi:10.1371/journal.pone.0058779.

[57] TANDY, C. R. The isovist method of landscape survey. *Methods of Landscape Analysis* (1967), 9–10.

[58] THOMAS, J. J., AND COOK, K. A. A visual analytics agenda. *IEEE Computer Graphics and Applications 26*, 1 (2006), 10–13. doi:10.1109/MCG.2006.5.

[59] TUFTE, E. R. *The Visual Display of Quantitative Information*, 2nd ed. ed. Graphics Press, Cheshire, Conn, 2001.

[60] TURNER, A. Angular analysis. In *Proc. 3rd International Symposium on Space Syntax* (Atlanta, Georgia, 2001), Georgia Institute of Technology, pp. 30.1–30.11.

[61] TURNER, A. From axial to road-centre lines: A new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: Planning and Design 34*, 3 (2007), 539–555. doi:10.1068/b32067.

[62] TURNER, A., DOXA, M., O'SULLIVAN, D., AND PENN, A. From isovists to visibility graphs: A methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design 28*, 1 (2001), 103–121. doi:10.1068/b2684.

[63] TURNER, A., AND PENN, A. Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. *Environment and Planning B: Planning and Design 29*, 4 (2002), 473–490. doi:10.1068/b12850.

[64] WEISMAN, J. Evaluating architectural legibility: Way-finding in the built environment. *Environment and Behavior 13*, 2 (1981), 189–204. doi:10.1177/0013916581132004.

[65] WIENER, J. M., FRANZ, G., ROSSMANITH, N., REICHELT, A., MALLOT, H. A., AND
     BÜLTHOFF, H. H. Isovist analysis captures properties of space relevant for locomotion
     and experience. *Perception 36*, 7 (2007), 1066–1083. doi:10.1068/p5587.

[66] WOLBERS, T., AND HEGARTY, M. What determines our navigational abilities? *Trends
     in Cognitive Sciences 14*, 3 (2010), 138–146. doi:10.1016/j.tics.2010.01.001.

[67] ZIMRING, C., AND DALTON, R. C. Linking objective measures of space to cognition
     and action. *Environment and Behavior 35*, 1 (2003), 3–16. doi:10.1177/0013916502238862.

Appendix A: Additional Visual Summaries

Batty 2001
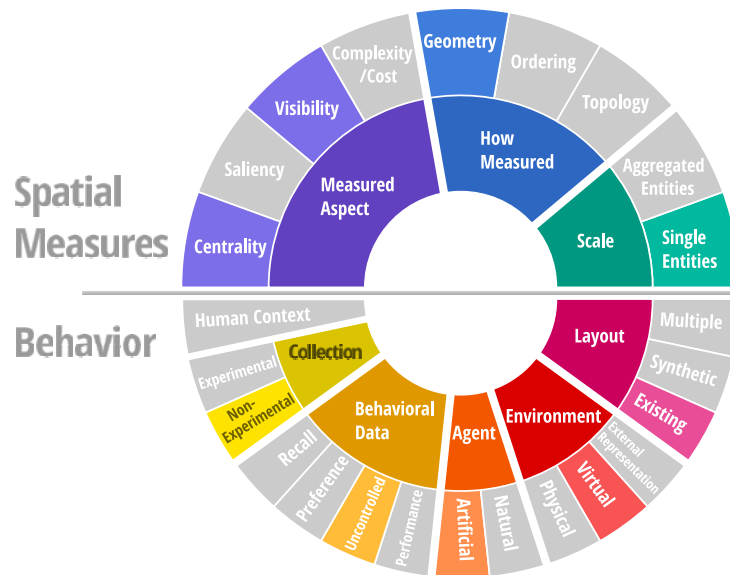
Chang 2002
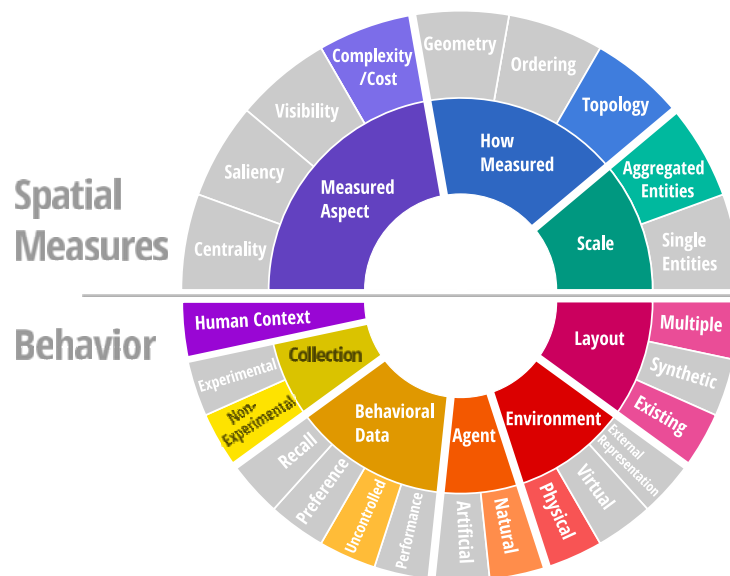
Dalton 2003
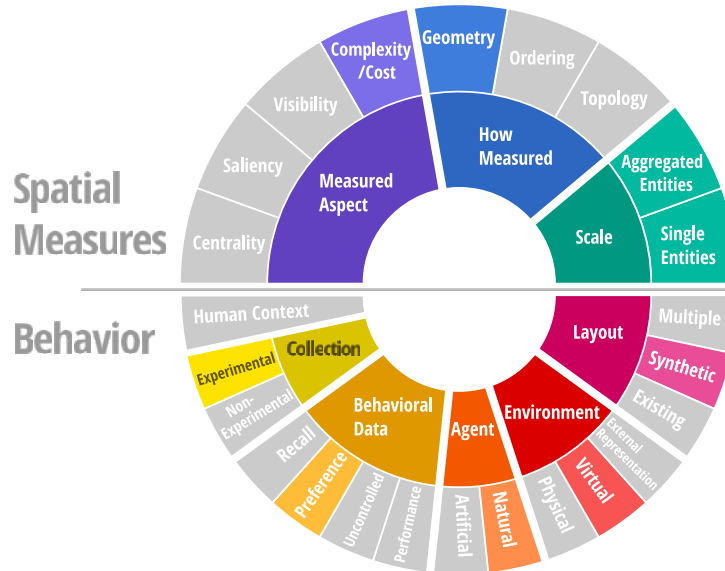
Hölscher et al. 2008
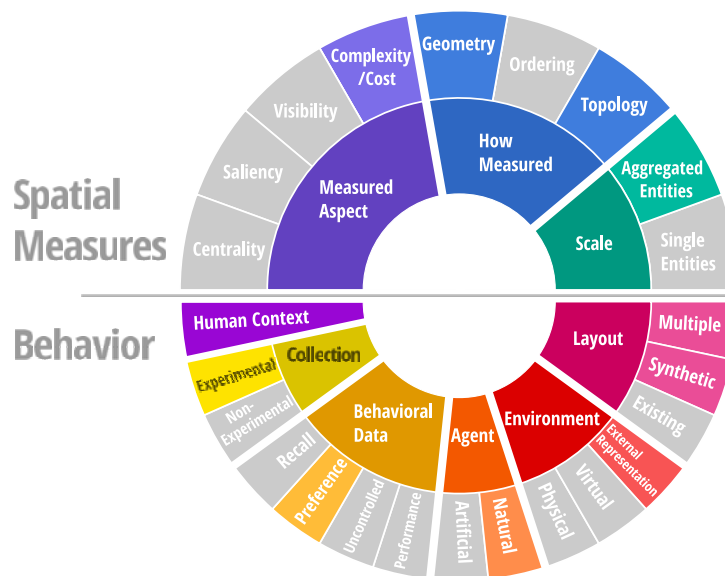
Hölscher et al. 2012
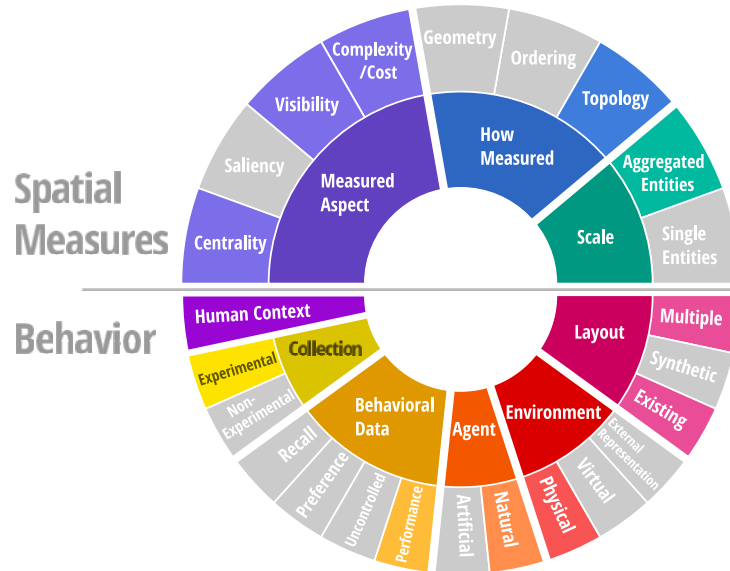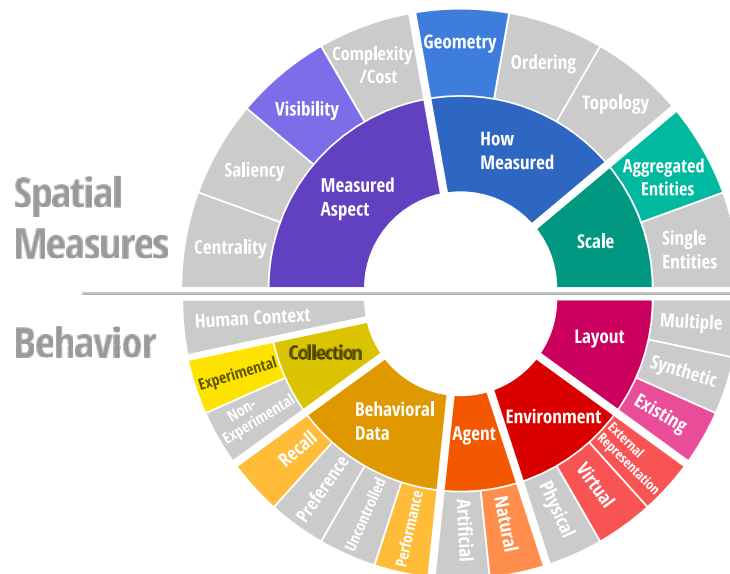

Jiang 2009
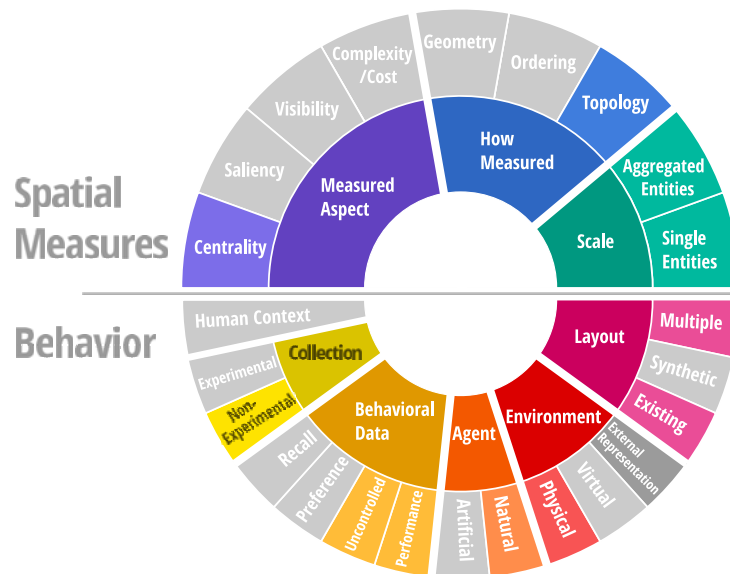
Koohsari et al. 2013
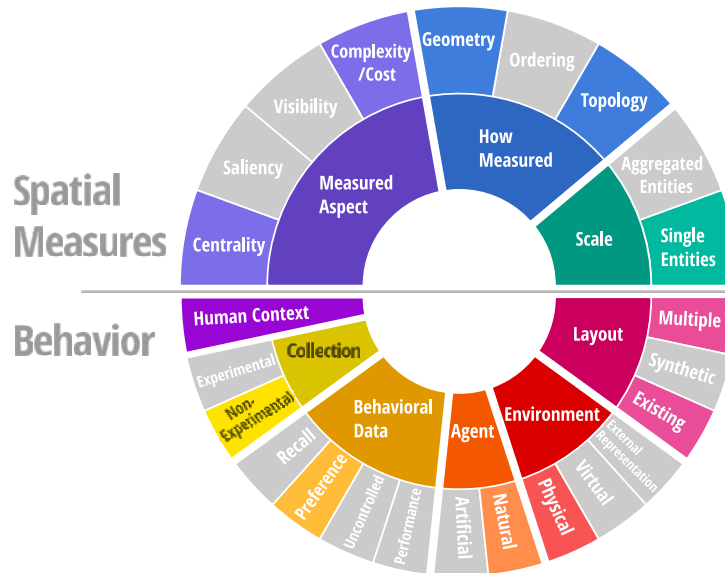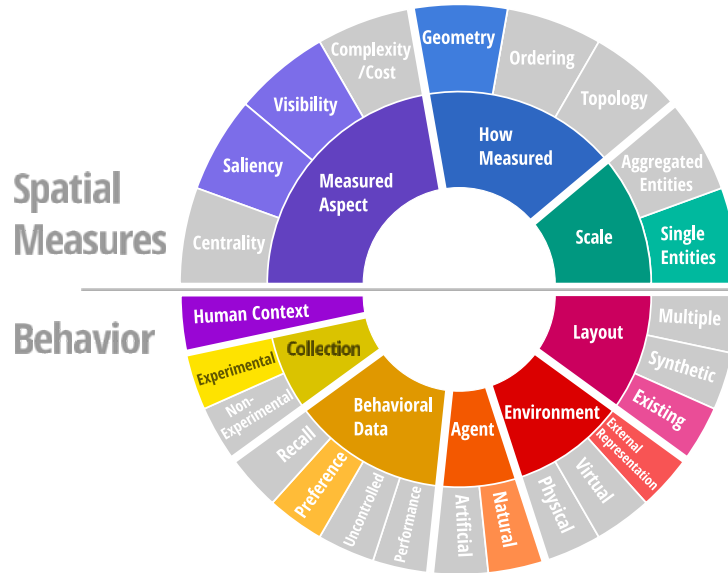

Li and Klippel 2014
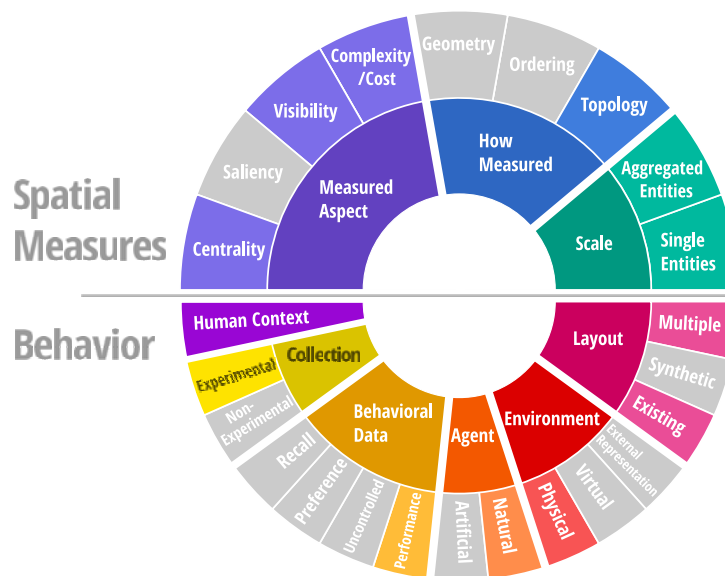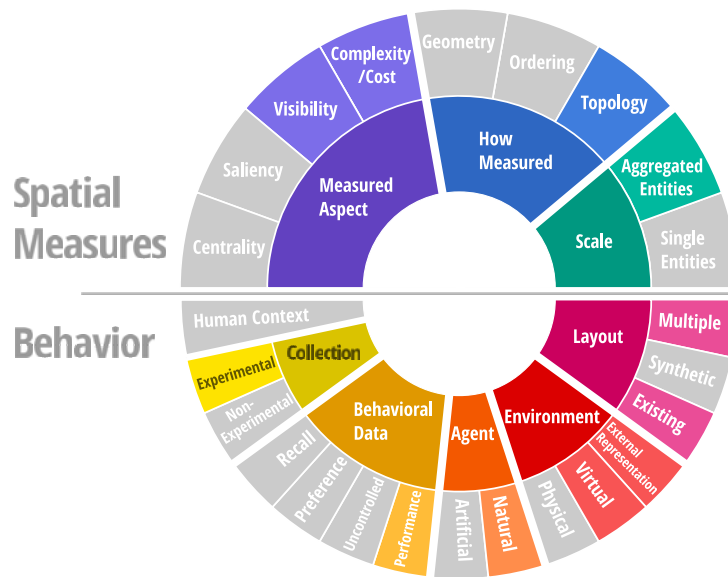
Lu and Zimring 2012
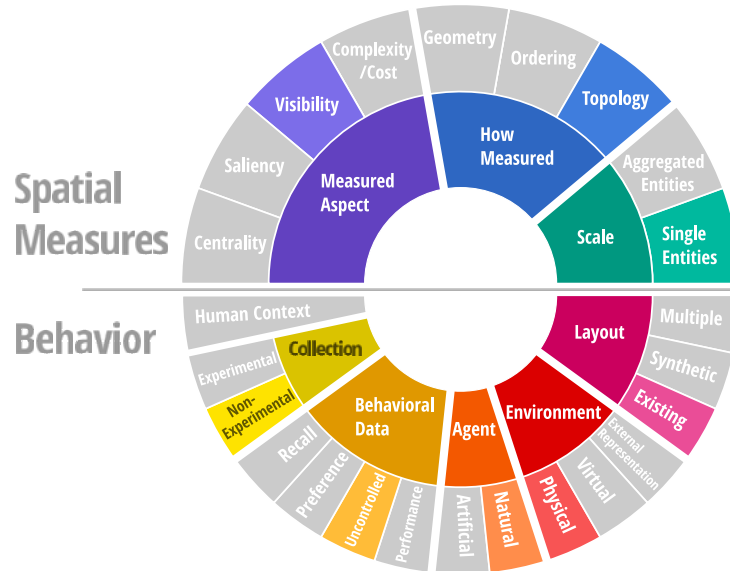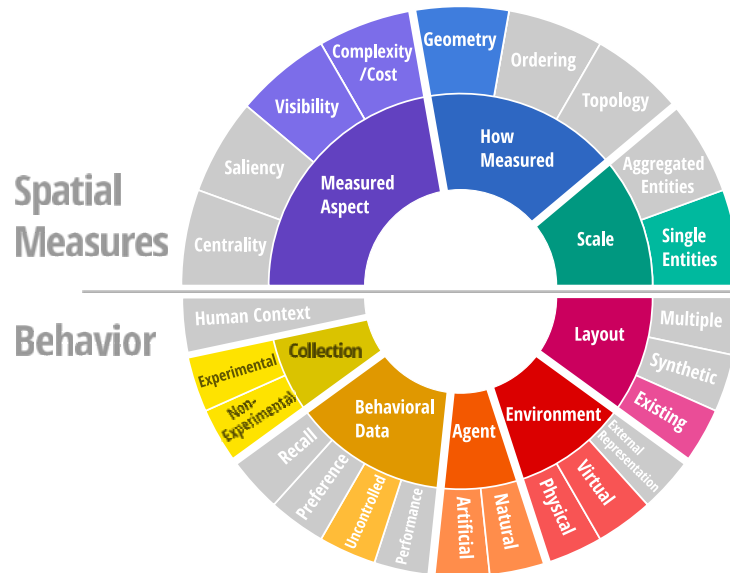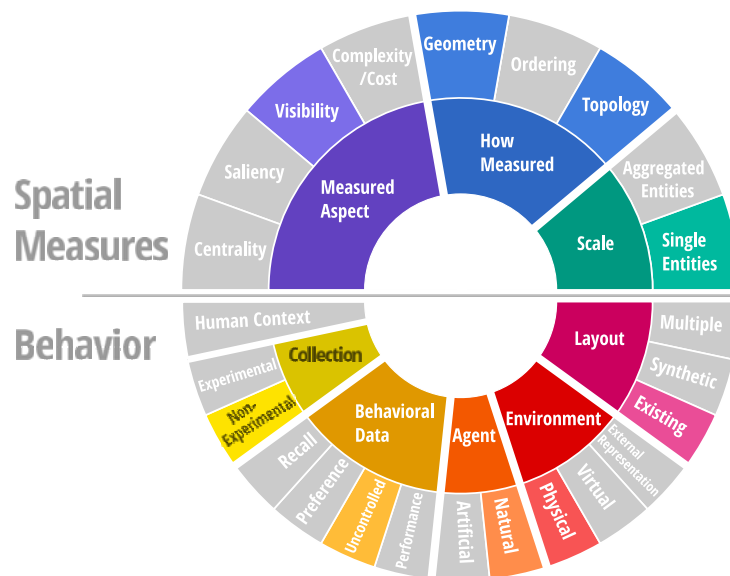

Meilinger et al. 2012

Nothegger et al. 2003
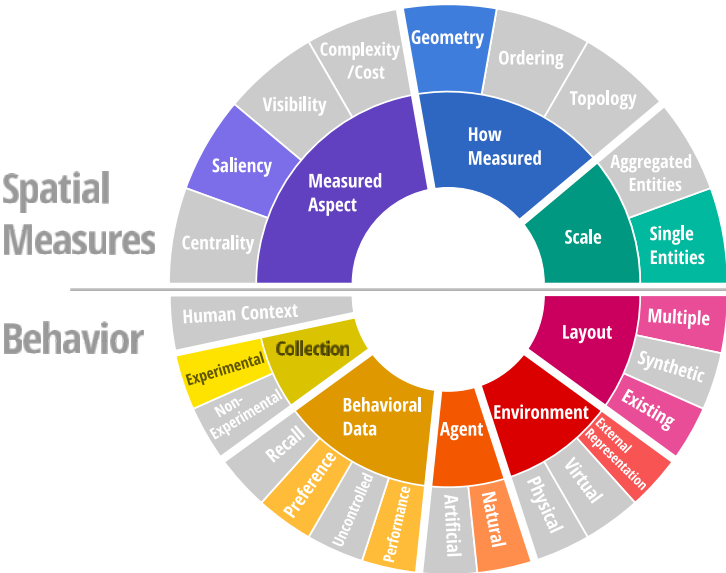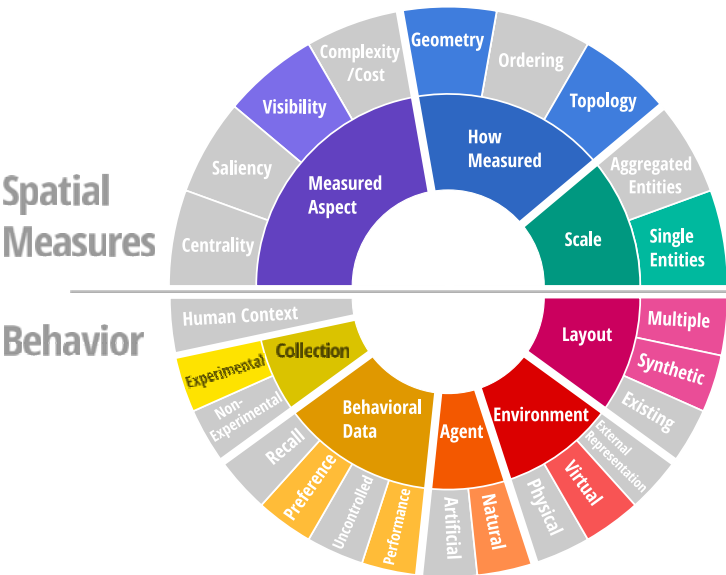

Omer and Goldblatt 2007

Turner and Penn 2002


Turner et al. 2001

Weisman 1981

Wiener et al. 2007

Appendix B: Category Definitions

Essential idea: This classification's central purpose is to help understand the state of the research literature regarding how quantifiable aspects of space influence human behavior in space. Accordingly, the classification is split into two major subsections: Spatial Measures and Behavior.
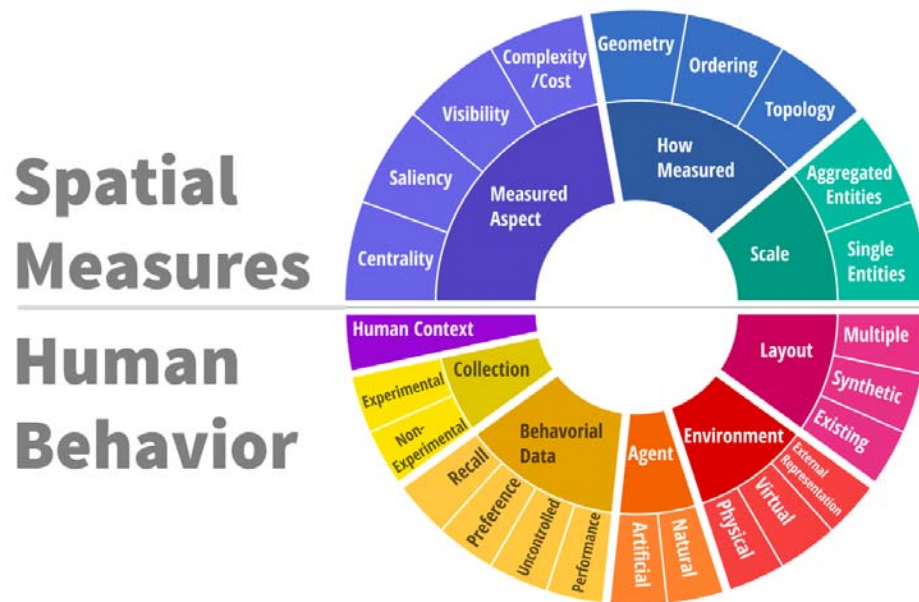
How to apply the classification scheme

- **General advice:** Be inclusive! Anything that you believe fits the study at hand should be marked in the classification. In particular, the following guidelines apply.

- **Measures within measures:** If authors measure some aspect by combining individual measures that cover other aspects, mark all these aspects in the classification (rather than just the ultimate result).

  Example: Nothegger, Winter, and Raubal (2004) utilize several environmental aspects, including visibility, to yield a landmark saliency measure. In that case, the paper is considered in this classification to include both saliency and visibility as the measured aspects.

- **Pre-existing data:** Research that analyzes data from previous studies or experiments and then also compares it with original data (data collected by the authors specifically for this research) is classified by taking the pre-existing study or studies into consideration together with the new contribution. That is, in the classification mark everything that applies either to the original data or the data from previous studies.

  Example: Turner and Penn (2002) correlate the results of an artificial agent model originated in that work to results of a previous study tracking the movement of hu-

man agents conducted by Hillier et al. (1996). In the classification, this would result in both artificial and human agent being marked, among others.

The classification scheme

In the explanations of the classification scheme's elements we will use the terms "space" and "entity" to refer to elements of the (geographic) area research has been conducted in. The terms are used according to the following definitions:

- **Space:** "Space" refers to an area as a whole. This may be on different scales (e.g., a room, a building, a campus, a city...), but it will usually be the encompassing area. "Space" may leave the internal structure of that area undefined.

- **Entity:** "Entity" refers to a specific, distinguishable part of a "space." This part may be a specific (geographic) object, such as a piece of furniture within a room, a room within a building, a building on a campus, or a district of a city. It can also be generic, referring to a not further specified location within a space.

Elements of the classification scheme

Spatial Measures

Measured aspect *The desired attribute of a space or entity, which measures are attempting to quantify.*

- **Centrality**- Importance of an entity based on its spatial relation with other entities. It can be established for structured (e.g., networks) and unstructured spaces. Centrality can be local (based on information of surrounding entities selected by some criteria) or global (based on information of all entities in a space).

  Example: Baran, Rodríguez, and Khattak (2008) compared space syntax measures of local and global centrality to walking behavior in different neighborhoods.

  Asami et al. (2003) sought to identify the most central locations within historic Istanbul. They compared centers found using the space syntax measure of integration (which derive values for centrality based on the topology of the street network) to centers identified by experts using a map, by the number of taxi bays in an area, and by the average number of stories of buildings in an area.

- **Saliency**- The distinctiveness or identifiability of an entity, relative to other entities.

  Example: Nothegger, Winter, and Raubal (2004) developed a model that uses attributes, such as color and facade area, to create a single measure of saliency for buildings.

- **Visibility**- The degree of visibility between one or more entities to or from another entity or set of entities.

  Example: Benedikt (1979) created a series of isovist measures to quantify the nature of visible space around a single point, and proposed methods to quantify visibility continuously across space.

- **Cost/Complexity**- How comprehensible the internal structure of a space is (i.e., how easy it is to understand a space) or the mental or physical difficulty of traversing a space (i.e., how easy it is to navigate and/or move through a space to a desired destination). This is often derived from a space's components.

  Example: O'Neill (1991) was interested in the effect of average intersection complexity on wayfinding task performance in relatively small scale spaces (sections of a library).

  Richter (2009) combined different aspects of a road intersection (such as number of branches and segment lengths) to define a measure for that intersection's cognitive complexity.

How measured?   *The actual mathematical structure used to derive values, outside the context of what aspect is being measured.*

- **Topology**- Measures of spatial relationships between entities in terms of connectedness or neighborhood without regard for geometric properties like metric distance, angle, size or shape.

  Example: Hölscher, Brösamle, and Vrachliotis (2012) identify areas of high centrality, using space syntax measures, which abstract a space into a graph. This graph abstraction allows for the purely topological relationships of the areas within a space to be quantified, such as how connected they are to other areas, without regard for geometric properties like metric distance.

- **Ordering**- Linear/circular order of a finite number of entities, without metric properties of distance or angle.

  Example: Richter (2007) uses circular ordering information of an intersection's branches and the position of a landmark object within that order to determine that landmark's location relative to a turn at the intersection (e.g., whether a landmark is located before or after the turn).

- **Geometry**- Quantitative (numeric) measuring of geometric properties such as shape, size, angle, and metric distance.

  Example: Examples include angular change (Turner 2001, Asami et al. 2003), and metric distance (Nagar and Tawfik 2007).

Scale   *This refers to the scale a measure operates on, i.e., whether it operates locally or globally. Scale **does not** describe the size (or resolution) of the space.*

- **Aggregated Entities**- Measures that summarize values measured for individual entities into (usually) a single measure for a larger unit (e.g., summarizing measures for individual intersections to a single measure for the whole route; calculating an average measure for a larger area from values for individual locations within that area).

  Example: O'Neill's (1991) global interconnection density (ICD), which summarizes the complexity of a space based on aggregating the number of decisions available at all intersections within a space, resulting in a single (average) value for that space.

- **Single Entity**- Measures which provide values for individual entities of a space and, thus, reveal differences between them.

  Example: Baran, Rodríguez, and Khattak (2008) use a space syntax graph abstraction (an axial line map) and centrality measure (integration) to yield a centrality value for each axial line, which defined segments of a pedestrian path network.

Behavior

Agent   *Beings whose behavior or actions are observed and compared to spatial measures.*

- **Natural**- Human agents, people used as test subjects.

  Example: O'Neill (1991) had graduate and undergraduate students perform wayfinding tasks.

- **Artificial**- Artificial agents whose actions are intended to approximate human behavior.

  Example: Turner and Penn (2002) created an agent-based model where the agents use visibility information to make movement decisions.

Behavioral data   *Type of data that captures the behavior of agents in some quantifiable way.*

- **Recall**- Measures how well a space and/or entities within that space are remembered. This is measured post-hoc, that is, after some task execution. It may be measured on different levels of spatial knowledge, such as landmark, route, or survey.

  Example: Omer and Goldblatt (2007) had human agents mark the location of landmarks on an incomplete map of a space in which they had previously performed wayfinding tasks.

- **Preference**- Indicated preference by agents given a set of choices.

  Example: Weisman (1981) had human agents judge a series of highly abstracted floor plans in terms of their level of general preference for the plan.

- **Uncontrolled**- Tracking an agent's decisions without a specific goal imposed upon the agents by the researchers. The researchers do not measure performance (of any kind), but simply observe behavior, such as how many agents pass by or enter a specific location.

  Example: Chang (2002) observed pedestrian movement by creating "gates" at particular points in a physical space and measured how many human agents passed through them, and secondly, he randomly selected human agents and tracked their movement through the space.

- **Performance**- In contrast to "recall" this measures the observable performance during task execution, such as time to completion, route optimization, or number of turning errors.

  Example: O'Neill (1991) used human agents in a wayfinding experiment, and measured their ability to find a predetermined destination in three ways: 1) Time elapsed, 2) Number of backtracks on route, 3) Number of wrong turns.

Collection   *How the behavioral data is gathered.*

- **Experimental**- Data recorded by researchers in an experimental study described in their paper. Researchers manipulate variables, such as the environment participants operate in, or the kind of information presented to participants, or the kind of participants, to observe causal relationships.

  Example: O'Neill (1991) conducted an original wayfinding experiment with human agents. Three portions of a university library were used as test spaces. They were selected to vary in value according to a complexity measure (InterConnected Density-ICD) based on intersection connectivity. Agents were given a task to locate a particular intersection within the environment, Their ability to find the intersection was tracked via the time it took them to find the specified location, how often they backtracked on their route, and how often they made suboptimal turns. The average performance in each space was then compared to the measured complexity according to ICD.

- **Non-Experimental**- Data that is collected without the controlled manipulation of variables, such as a survey, an observational study, or census data.

  Example: Koohsari et al. (2013) mailed surveys that asked human agents to report how often they walked to nearby "public open spaces" (i.e., parks) and how much time they spent doing so.

Environment   *The way the environment/space is experienced in the study.*

- **Physical**- An actual physical space as it exists in the real world.

  Example: O'Neill (1991) had subjects perform tasks in a university library building.

- **Virtual**- A computer-simulated space which emulates the rules of physical reality, such as a digital 3D model ("virtual environment").

  Example: Meilinger, Franz, and Bülthoff (2012) had human agents experience a 3D simulation of a town through a 220° semi-cylindrical screen.

- **External Representation**- Studies that utilize an (often static) representation of a space, such as a floorplan, map, or series of photographs, in addition to or instead of having agents interact in or move within a real or simulated environment.

  Example: Asami et al. (2003) had designated experts select local centers in Istanbul using a map of the city.

  O'Neill (1991) had human agents first experience an environment through the use of a series of photographs in order to familiarize them with the environment before encountering it directly.

Layout   *The type of spatial structure or arrangement of the environment.*

- **Existing**- A spatial layout found in the real world that could be encountered in life.

  Example: Hölscher, Brösamle, and Vrachliotis (2008) had human agents perform tasks within a pre-existing building (a conference center).

Meilinger, Franz, and Bülthoff (2012) had human agents experience a virtual model of the town of Tübingen, Germany.

- **Synthetic**- A layout that is designed specifically to test behavioral responses to environmental aspects, such as a maze or idealized regular street network.

  Example: Dalton (2003) created a virtual, idealized urban form consisting of streets of all the same length, but a variety of intersection types.

- **Multiple**- The study compares multiple layouts, i.e., multiple different spaces (or multiple distinct areas within a space).

  Example: O'Neill (1991) compared three different areas within a library which varied in measured complexity, in order to validate the measure against human behavior.

Human Context   *A study also tests for characteristics of human agents, such as familiarity, expertise, sex or individual differences.*

  Example: Nothegger, Winter, and Raubal (2004) compared landmarks at street intersections as defined by a saliency algorithm to those selected by human agents with varying degrees of reported familiarity with the space.