August 2018 To appear in *Spatial Cognition and Computation*

Geometric Problem Solving with Strings and Pins

Christian Freksa FREKSA@UNI-BREMEN.DE
Thomas Barkowsky BARKOWSKY@UNI-BREMEN.DE
Zoe Falomir ZFALOMIR@UNI-BREMEN.DE
Jasper van de Ven VANDEVEN@UNI-BREMEN.DE
University of Bremen, Bremen Spatial Cognition Center, Bibliothekstr. 1, 28359 Bremen, Germany

Abstract

Humans solve spatial and abstract problems more easily if these can be visualized and/or physically manipulated. We analyze the domain of geometric problem solving from a cognitive perspective and identify several levels of domain abstraction that interact in the problem solving process. We discuss the roles of physical manifestations of spatial configurations, their manipulation, and their perception for understanding problem solving processes. We propose an extension of the classical problem solving repertoire of constructive geometry to approach certain problems more directly than under the compass-and-straightedge paradigm. Specifically, we introduce strings and pins as helpful metaphors for a generalization of the constructive geometry approach. We present classical problems from spatial problem solving to illustrate the 'strings and pins' paradigm. Three case studies are discussed: strings-and-pins solutions to (i) the ellipse construction problem; (ii) the shortest path problem; and (iii) the angle trisection problem. Comparisons to formal solutions are drawn. Differences and similarities between the compass-and-straightedge paradigm and the strings-and-pins paradigm are analyzed. Features and limitations of constructive and depictive geometry as well as implications for computational approaches are discussed. The strings-and-pins domain is shown to be more general and less restrictive than the compass-and-straightedge domain.

1. Understanding through seeing, doing, and abstract thinking

Most people 'grasp' spatial problems and their solutions more easily if they can visualize them or perform physical actions and observe their effects than if they rely purely on textual or formal descriptions (Kozhevnikov *et al.* 2007; Scaife and Rogers 1997, Brown 2008). This is an indication that spatial cognition (including perception and action) can support insight into problems in certain situations more effectively than (textual or formal) linear descriptions. On the other hand, linear descriptions are easier to handle analytically by humans or by computer programs. Therefore, to understand the potential of (natural or artificial) cognitive systems for grasping and solving problems, it appears worthwhile to investigate spatial capabilities and spatial challenges they face. This can be done by studying operations and processes that are performed directly in the spatial domain (Freksa 2015).

The connection between formal analysis and spatial construction is widely known through the relation between formal and depictive[1] geometry (Euclid 1926; Larkin and Simon 1987). Proofs in depictive geometry can be formally described. However, formal proofs generally are structurally not fully equivalent to the corresponding depictive proofs. The reason for the lack of equivalence is that the spatial medium of a diagram (typically a piece of paper or a board) *implicitly* contributes relevant spatial relationships and dependencies that must be *explicitly* described and handled in the formal approach. In other words, spatial relations in a diagram are an essential part of the depictive representation whereas spatial relations between formulas on the paper are not a part of the formal representation. Thus, although both the depictive and the formal representation may be presented on the same medium, in the depictive case the structure of the medium is a relevant integral part of the representation, while in the formal case it is not.

Often formal proofs are found with the help of insights resulting from visuospatial understanding, and could not have been found as easily on a purely formal basis. This phenomenon is not restricted to spatial problems; trained theoreticians advise their students to first try to understand abstract problems through diagrammatic depictions, i.e. spatializing the problems, before tackling them formally, even though the depiction may be more constrained than the abstract problem to be solved. This is in line with the long history of visual thinking in mathematics; see for example (Polya 1945; Giaquinto 2007).

1.1 Geometric reasoning with diagrams vs. formal reasoning

A nice demonstration of the difference between diagrammatic and formal reasoning is presented by Bernays (1976), who compares Euclid's geometric approach with Hilbert's abstract theory of geometry (Mumma 2012). In Euclid's approach, an interpretation of geometric objects is necessary to understand a proof, whereas Hilbert's proofs only depend on the logic form (i.e. a geometric interpretation – although possible – is not presumed). Mumma also points out that, while Euclid's approach is constructive, Hilbert's formalization is existential.

One example given in Mumma (2012) is the concept of 'opposite' that expresses that two points are located on opposite sides of a line. Simplified, in Euler's view, by construction, two points are on opposite sides of a line if you draw them as such (Fig. 1 left). In Hilbert's case, this is not sufficient. Hilbert requires the construction of an auxiliary line p_1p_2 and the existence of some point (x), which needs to be located on the line section between p_1 and p_2 and on (l). Formally this is given by $OppSide(p_1, p_2, l) \leftrightarrow \exists\ point\ x\ B(p_1\ x\ p_2) \wedge x\ lies\ on\ l$, with the ternary relation B denoting that x is positioned between p_1 and p_2 (Fig. 1 right).

The proofs by Euclid have been investigated by Manders (2008) to explicate that they make use of two types of structures: the diagram and the accompanying text. Furthermore, Manders distinguishes between *exact* and *co-exact* properties. Considering inaccuracies in manually drawn diagrams, for example an imperfectly drawn straight line, exact features are those which change in the presence of smallest variations; co-exact features are those which remain stable in the presence of these inaccuracies. This is related to differences in quantitative and qualitative reasoning. Quantitative descriptions are sensitive to arbitrary small changes whereas qualitative

[1] We use the term *depictive geometry* instead of *descriptive geometry* in order to avoid confusion with linguistic and formal ways of describing geometric relations.

descriptions can focus on distinctions between objects that make an important and relevant difference with respect to a given problem (Kuipers, 1994). In quantitative descriptions, reasoning is performed on numbers, where small changes will directly influence the result; in comparison, in qualitative descriptions, reasoning is performed on a symbolic categorical level, where small deviations may not make a difference.

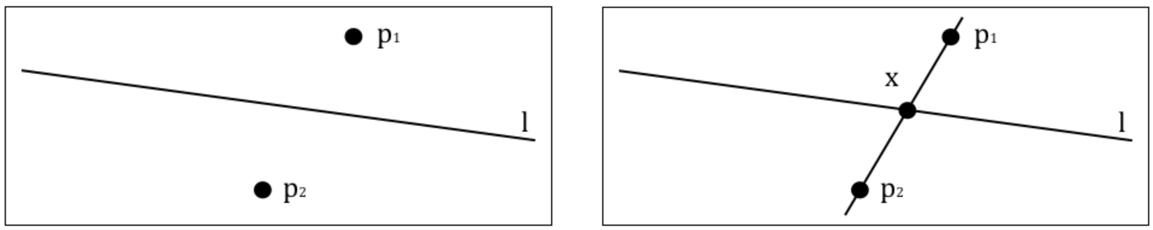

Figure 1. Euclid's understanding of 'opposite' (left) and Hilbert's abstract definition by means of an auxiliary line and an intersection point (right)

1.2 Comparison between descriptive and depictive representations

The contrast between descriptive and depictive representations can be observed by comparing a 2-column table, showing numbers and their squares, with a graphic depiction of those numbers, relating the magnitudes of the numbers spatially and visualizing them graphically. In order to understand the table (Fig. 2 left), we must mentally represent the numbers and relate their values to one another abstractly or by mental imagery (Finke 1989). In the graph (Fig. 2 right), the values are associated through spatial relations; we can visually read off relations between the values by comparing distances; we also can directly sense slopes corresponding to relative differences of numbers in the two columns on the left. Although the concrete spatial distance is not identical to the relation between the abstract numbers, we understand the values and their relations more easily through spatial reification than through number-theoretic arguments. For an enlightening discussion on how pictures can be superior not only for visualizing relationships but also for proving theorems, see Chapter 3 of Brown (2008).

x	$y = x^2$
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9

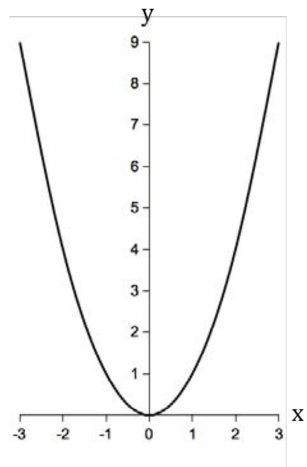

Figure 2. Descriptive (left) and depictive (right) representation of abstract information. In the descriptive representation, spatial relations in the table only relate x-values to y-values, while in the depictive representation magnitudes and variations also are spatially related and visually accessible.

It is well known that different kinds of representation afford different kinds of reasoning (Sloman 1985). Similarly, different kinds of tools or metaphors afford different kinds of conceptualization. In the present paper, we demonstrate the roles of spatial structures and of physical metaphors in the domain of classical Euclidean geometry.

1.3 Contributions of this paper

This paper analyzes geometric problem solving from a cognitive science perspective. It discusses spatial problems well known from classical geometry and computer science and scrutinizes the concepts, metaphors, tools, and approaches that have been employed to solve them. We identify different levels of abstraction that play a role in geometric problem solving. We propose a generalization of classical concepts for geometric problem solving. Using different metaphors for conceptualizing tools for geometric constructions we demonstrate how a larger set of problems can be tackled in the methodological framework of classical geometry.

We address the relation between abstract spatial concepts and concrete manifestations in the physical realm. We discuss how abstract concepts can be applied to the real world and how real-world spatial configurations can be used for abstract geometric problem solving entirely within a spatial substrate and its conceptualization. In this way we extend the range of constructive spatial problem solving. For an enlarged class of spatial problems we can avoid the need of formalizing spatial configurations and the associated problems and risks of distortion of crucial properties.

Specific examples are given to illustrate the approach. The examples also serve to discuss the roles of physical manifestations and metaphors in constructive geometry – or more generally – in problem solving. The conceptual beauty and pureness of feature-preserving spatial transformations is shown. Advantages of structure-preserving representations are elicited. Finally, conclusions are presented.

2. Levels of abstraction in geometric construction and reasoning

From a cognitive perspective on spatial problem solving and geometry, we can distinguish at least five domains involved in geometric construction and reasoning that are related to one another:

1. Abstract geometric concepts *(point, line, circle, …)*
2. Conceptual tools to manipulate geometric concepts *(compass, straightedge, string, pin)*
3. Physical tools that the conceptual tools refer to *(drawing compass, ruler, cord, peg)*
4. Physical / conceptual agents that make the tools work (e.g. human or robot)
5. Physical manifestation of geometric concepts (image of a circle, circular flowerbed)

Some or all of these domains are involved in geometric problem solving processes – or, more generally, in spatial problem solving processes. We will sketch them briefly in the following subsections.

2.1 Abstract geometric concepts

Abstract concepts in geometry include the primitive concepts *point* and *line* and additional concepts like *circle* or *triangle* that can be constructed by mathematical operations or by diagrammatic construction. This is the level that the conceptual and mathematical domains of geometry are concerned with and where the semantics of geometric constructions is defined.

2.2 Conceptual tools to manipulate geometric concepts

These are the tools mathematicians use to represent and solve geometric problems. They include (a) informal descriptions for verbal communication, e.g., *the set of all points in a plane that are at a given distance from a given point* to characterize the concept of a circle; (b) algebraic formulas to define and manipulate geometric concepts, e.g. the algebraic equation $x^2 + y^2 = r^2$ to define a circle with radius r centered at the origin (0, 0); or (c) mental images or metaphors of physical tools to construct representations of geometric concepts that fulfill the required criteria, e.g. the notions of a *compass* and a *straightedge* in geometry.

2.3 Physical tools to manipulate geometric concepts

The conceptual tools mathematicians employ to describe the construction of geometric entities often refer to perceivable physical objects. Frequently we use the same labels for the conceptual entities as for the physical ones; this makes it difficult to distinguish the physical and the conceptual levels. In this paper we use different labels to distinguish the conceptual and physical levels, e.g. straightedge vs. ruler; string vs. cord; pin vs. peg; and compass vs. drawing compass, as *conceptual and physical tools have different cognitive roles*. Examples of physical tools are drawing compasses (with needles on one side and lead pens on the other side), rulers with pencils to be applied to them, or cords that gardeners may use to mark straight lines or to construct shapes like circles or (other) ellipses.

2.4 Physical / conceptual agents that make the tools work

For physical tools it is quite obvious that they require an agent to become functional: drawing compasses need to be moved in order to depict a circle, rulers needs to be moved to align with certain points, etc. But even if we do not employ physical tools but merely imagine the physical actions for creating geometric constructions in our minds, we need to imagine the operations we would carry out with physical tools, and this requires some mental agency. Alternatively, more abstract thinkers, who do not employ mental imagery for solving geometric problems, but who mentally manipulate formulas, also require an agency that controls the manipulation of the formulas.

The agency that controls geometric constructions and geometric reasoning presumably employs meta-knowledge derived from learning and experience in order to solve geometric problems. We need to understand the nature of this knowledge if we want to enable artificial cognitive systems to solve geometric problems autonomously.

Physical constructions of geometric configurations only are useful for geometric problem solving if there is an agency capable of perceiving them. The perception of spatial configurations can take place visually or haptically. Imagination can be viewed as a kind of mental perception. Perception including interpretation requires active processes that control the attention and are able to make sense of the perceived patterns. Thus, we also need to understand the nature of perception and interpretation in order to enable artificial cognitive systems to employ diagrams or other constructions for geometric problem solving.

2.5 Physical manifestation of geometric concepts

Mental images and metaphors that humans use to reason about geometric concepts and to characterize them refer to perceivable physical entities. Examples of physical depictions of geometric concepts are pencil drawings of points, lines, and circles on a sheet of paper. These physically manifested shapes share some, but not all properties of the corresponding abstract concepts. Specifically, they disagree in material aspects (consisting of lead) and geometric aspects (having non-zero width). However, it is exactly those aspects that make these physical entities cognitively useful: the physical properties and deviations from the ideal geometric concepts are responsible for making these concepts perceivable to humans and thus they support our imagination capabilities. Whether the geometric properties of imagined depictions correspond more closely to the physical properties or to the abstract ideals might be an interesting question to investigate.

3. Spatial media for spatial problem solving

Great discoveries and proofs have been made in geometry for over two thousand years through combining the concept of a circle with the concept of a straight line. These concepts are easy to comprehend as they can be made perceptually accessible through a drawing compass or through a piece of cord that is fixed on one end and can be pulled around the fixation point (in the case of a circle) and through a straight edge, a taut string, or a line of sight (in the case of a straight line).

Classical geometry has investigated the range of geometric constructions that can be performed if a circle is constructed by a compass and a straight line with a straightedge in an idealized plane geometric world. An amazing range of geometric constructions can be performed by compass and straightedge alone. However, certain problems that appear to be straightforward have been proven to be unsolvable by a compass-and-straightedge construction.

From a cognitive science perspective it may be interesting to note that although constructive geometry appears more tangible than other fields of mathematics, it takes place entirely in the conceptual domain of mathematics. ‘Compass’ and ‘straightedge’ correspond to abstract notions connected with abstractly defined properties. The two notions merely serve as metaphors to support our imagination. This becomes apparent when we note that the tools for geometric constructions do not mention paper, pencil, or other media through which to apply the construction; accordingly, neither the compass nor the straightedge need to leave a trace in form of a pencil line. For geometric reasoning it is sufficient to know that intersections between lines can be constructed and therefore exist and that we can use them as concepts for mental constructions. Paper and pencil serve to support our imagination and reasoning processes through generating a perceivable analogy of the conceptual structures and operations and to help us to communicate geometric concepts to fellow human beings.

Stimulated by the shortest path finding procedure of pulling apart the starting and the destination nodes of a street network represented by an undirected graph constructed out of strings (Minty 1957; Dreyfus and Haugeland 1974; Freksa 2015) – see Section 6.2, we set out to investigate whether the concept of a ‘string’ could be employed as a metaphor to extend the range of spatial problems to be solved in the spirit of Euclidean conceptualizations. We define a ***string*** to be an idealized deformable ***cord*** of unlimited length and negligible width.

Strings are conceptually appealing for spatial problem solving for at least two reasons: (a) a string can be deformed in useful ways in which we cannot deform the straightedge in Euclidean geometry (e.g., Cabalar and Santos 2011); and (b) a string can be pulled taut to form a straight line; it then is equivalent to the straightedge in Euclidean geometry.

We add the notion of a ***pin***. This notion is inspired by the pin-shaped end of a compass. A pin identifies a point location in the geometric plane or on a string. Two pins on a taut string define a ***line section***.

A string and two pins instantiate the concept of a compass. A taut string with two pins, one of which is fixed at a given location, define a ***circle***. Thus, we introduced new metaphors which by construction instantiate variants of the compass and straightedge that are equally capable of performing Euclidean circle and line constructions. We can use the notion *taut string with two pins* instead of *compass* and the notion *taut string* instead of *straightedge* for all constructions in Euclidean geometry.

4. Extending the tool set for geometric constructions

Manipulation and perception of spatial configurations are essential for generating geometric constructions. We will explore whether we can extend the set of useful constructions in order to enlarge the range of spatial problems that (a) maintain spatial properties implicitly by exploiting intrinsic spatial structures of flat 2-dimensional surfaces; and (b) make use of the rigor and beauty of constructive geometry.

To this end, we generalize the compass-and-straightedge metaphor. We permit strings to form arbitrarily shaped lines. We conceive of a string section as having constant length, independent of its shape. The length of a string section is equal to the distance between the two section-defining pins in the taut string section. The idea is that in addition to classical geometric constructions this generalization will allow us to reason about the lengths of arbitrarily shaped lines.

It may appear awkward to include entities with undefined shape into the set of geometric objects; however, in constructive geometry, knowledge of the objects tends to be rather limited: sometimes we know the general shapes of the entities drawn due to construction (line section, arc of a circle), but we have little knowledge of most locations on these curves, unless – by construction – we define boundary points or intersections. In the strings-and-pins domain, we have entities whose lengths are given and remain constant under transformations we impose on them.

Pins can be used in several ways as fixed or floating entities: (a) they may have an absolute location (pinned to the flat reference world, e.g. a sheet of paper like the needle of a drawing compass); (b) they may define locations relative to a location on a string (pinned to that location of the string); (c) they may (dynamically) mark a variable location on a string (e.g. dividing a string section into two parts, whose sum is constant). Pins also can serve combinations of these uses, i.e. they may fix locations of lines to locations on the reference plane (like a pencil mark on a sheet of paper); they may collocate locations on two or more strings (either fixed to an absolute location or only relative to one another). The use of these operations will be illustrated in Section 6 (Examples).

As we discussed in Section 2, geometric construction and reasoning involves various levels of abstraction. As long as we stay purely on the abstract conceptual level, we do not have to worry about representational issues; but if we want to apply geometric concepts to real-world situations or if we want to use real-world configurations for geometric reasoning, we must address the relations between the abstract concepts and the real objects (Freksa 1997). These relations are outside the theory of geometry, but they become relevant when we look at problem solving from a cognitive perspective.

Representational issues are even more obvious when we consider formal approaches to spatial problem solving: how can we formalize a real-world problem and how do we interpret a formal solution to such a problem in terms of real-world configurations? These questions are currently outside the domain of AI systems and cannot be handled automatically. Human experts and their cognitive abilities are required to generate and to interpret formal representations in order to establish a correspondence between the abstract world of computation and the real world of physical entities. In the following two subsections we will address the mapping between the abstract domain and the real world.

4.1 Applying abstract concepts in the real world

Geometric operations are defined in the domain of abstract concepts. If we want to use geometry to solve real-world problems we need to map abstract results into the real-world domain. This mapping is outside the domain of pure geometry; for humans this usually does not cause problems as the domain of geometry is considered to be an idealized geometrically similar copy of the corresponding real-world domain, and human perception and cognition can easily identify corresponding structures in the two domains. The results of abstract geometric constructions can be applied one-to-one by an agent who has access to both, the geometric and the real-world domain.

Gardeners and other real-world workers overcome the representation problem of an abstract to real mediation process by applying concepts of geometry directly in the real world. For example, a gardener who wants to create a circular flowerbed around a tree may use the stem of the tree as center of the circle and a taut cord sliding around the tree to construct the circle in the real domain. The radial distance can be marked on the cord by a pin-like object (a *peg*) to leave a trace of the circular boundary. This example illustrates how close we can get in applying abstract concepts to real environments; many abstract representations involve more complex interpretation processes.

4.2 Making real-world entities accessible to abstract geometric tools

The other direction is more critical: How do we get real-world spatial configurations into the abstract domain of geometry? High school geometry avoids this issue by describing real-world problems abstractly before solving a problem, e.g. 'a flowerbed has triangular shape; its edges measure 3m, 4m, and 5m, respectively; …'; here a geometric shape and abstract length values are postulated. These postulates have been generated outside the geometric domain by a human observer, who mediates between the real world and the abstract domain. Strictly speaking, the correctness of this mediation process cannot be validated, as it is outside the theory of geometry.

In Section 4.1 we have seen how we can apply abstract geometric concepts in the real world; but how can we make real-world entities accessible to the full range of geometric operations that may be feasible? Consider the following situation: we look at a road map to identify suitable alternatives for the most direct route, which is currently blocked. Of course we could try to measure the lengths of road sections, enter them as abstract values into a computer-adept graph structure, and apply Dijkstra's algorithm (Dijkstra 1959) to find the shortest alternative route. But usually we will not do this – in the interest of not losing even more time than due to the blocked route. Most likely, we will easily identify a small number of candidate substitution routes to which we will apply visual or mental length comparison of some sort in order to identify the possibly shortest route.

We now would like to put the cognitive approach to identifying a shortest path on a scientifically more solid basis. The question we want to address here is: 'how do we get the relevant route information from the map into a shortest path identifying mechanism'. In Section 6.2 we will then address the shortest path identifying mechanism itself.

Note that we are not interested in the length of the shortest path; we are only asking for a shortest alternative. Therefore we would like to avoid the procedure of measuring lengths of road sections and of dealing with absolute length value abstractions. Besides, the measuring procedure could only be an error-prone approximation.

In the real world, routes are manifested through road connections on the surface of the earth. In order to perceptually deal with and reason about roads, humans typically exploit a structure-preserving *mild abstraction* (Freksa *et al.* 2018) that maps a real-size route network spatially into the size range of human perception: they use a scaled map of the environment. A linearly scaled map preserves relative lengths; this means that the shortest route on the map will correspond to the shortest route in the real world. Imagine for a moment that roads on the map are manifested by cords that are connected at road intersections; we then have a cord network that we can manipulate in similar ways as the gardener who constructs the circular flowerbed around the tree, while structurally maintaining crucial geometric properties.

5. Abstract concepts, physical metaphors, and formal representations

The beauty of constructive geometry is in the validity of the result as a function of the postulated initial conditions for the construction. Strictly speaking, these postulated conditions only hold in the conceptual domain. By using metaphorical language such as 'compass', 'straightedge, 'string', and 'pin' to characterize conceptual tools, we map these abstract conditions to mental images of physical tools that maintain the conditions. If we take the metaphors literally, we can approximate the conceptual conditions to the extent to which the suggested physical tools fulfill these conditions. For example, the needle of a drawing compass approximates a geometric point location; the distance between the needle tip and the pencil tip approximates a constant distance; the pencil line approximates an infinitesimally thin line; etc. Thus, physical interpretations of abstract geometric configurations approximate the properties of these configurations, for example we obtain approximate circles, approximate tangents, etc.

Abstract geometric reasoning about real-world entities also is valid only to the extent to which we are able to map real-world properties correctly to abstract properties. In particular, if we formalize physical spatial dimensions by measuring distances, angles, etc. we will introduce imprecision, and the formal representation will be an approximation of the represented configuration. Abstract geometric reasoning can only be valid with respect to the real-world domain to the extent to which the operations in the abstract domain are benign with respect to corresponding relations in the real world, that is, allow the reconstruction of the relevant relations in the real world. A theory of abstract reasoning cannot guarantee such benign correspondences, as real-world properties are outside these theories. A suitable representation theory would be required that includes the real-world domain and its properties, the abstract formal domain and its properties, and the mappings between the two domains (c.f. Palmer 1978).

The formalization of problems is carried out on the basis of intuition rather than on the basis of formal theories (as formal theories can be applied only to formal representations). One approach to gain trust into the suitability / benignity / fidelity of problem formalizations would be to employ several structurally different formalizations of the same real-world domain. If these different formalizations would yield different abstract results under their respective different formal operations, we would know that some of them are inadequate; if, however, we would obtain identical results, we might gain confidence in their correctness.

An alternative approach to gain confidence in the appropriateness of a representation is to keep the structure of the representation as close to the structure of the problem domain as possible. In this way, the operations carried out in the representation domain may be identical to or closely correspond to the operations in the problem domain and the risk of misrepresentation due to structural transformation could be reduced. In other words: we use an analogical or even identical problem representation to solve the problem at hand. We call this approach to problem representation *mild abstraction* (Freksa *et al.* 2018).

6. Example constructions with strings and pins

We will now consider three well known spatial problems concerning distances and angles that go beyond geometric compass-and-straightedge constructions. In Section 6.1 we present the ellipse construction problem. This is a formal problem ("the points whose sum of distances to the focal points is constant"), the gardener's solution is a structure-preserving approximation in the real-world domain. In Section 6.2 we present the shortest route problem. Here, the problem domain is given by a spatial configuration in the real-world domain rather than by a formal representation. In Section 6.3 we present a string-and-pin-based solution to the formal angle trisection problem, a problem that has been proven to be unsolvable by compass and straightedge constructions.

6.1 Constructing an ellipse

In planimetry, an **ellipse** is a curve surrounding two focal points such that the sum of the distances to the two focal points is equal for every point on the curve. As such, an ellipse is a generalization of a circle, where both focal points are collocated. The geometric construction of an ellipse exceeds the capabilities of a compass, as one leg of a compass is confined to a single location while a general ellipse requires reference to two focal points.

A common method to constructing an ellipse (abstract / physical gardener's method) uses a string / cord that is formed to a closed loop which surrounds two pins / pegs that are positioned at the two focal points. The third pin / peg pulls the string taut to shape the string / cord to form a triangle whose vertices are located at the three pins / pegs. The pins / pegs at the focal points are fixed while the third pin / peg floats around them keeping the string / cord loop taut. We thus generate triangles whose base line corresponds to the line section connecting the two focal points and whose other two edges sum up to a constant length. The line along which the floating pin / peg moves thus depicts an ellipse (Fig. 3).

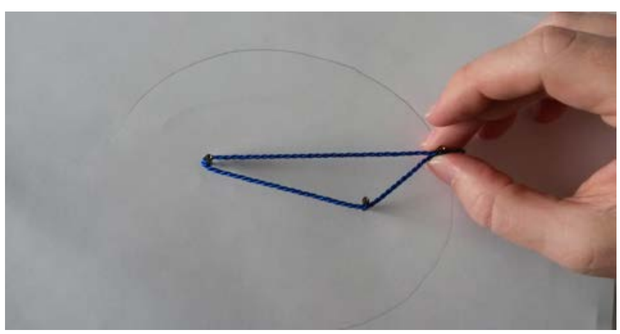

Figure 3. Constructing an ellipse physically, using three pegs and a cord.

6.2 Constructing a shortest path in a route network

The shortest path problem addresses the problem of identifying the shortest path between two specified nodes in a connected route network. Computational graph search approaches for solving the shortest path problem are well known, e.g. Dijkstra (1959). These approaches require quantified knowledge about the lengths of route sections in the graph. Here we will present a constructive approach which transforms a physical or conceptual spatial network within the spatial domain. The approach directly exploits affordances and constraints of spatial relations without reasoning about them. It transforms a given spatial configuration into a length-preserving new configuration that makes it easy to identify the shortest connection between the two nodes. The approach was presented by Minty (1957) and discussed by Dreyfus and Haugeland (1974) in their criticism of computational models of the mind.

The strings-and-pins approach to solving the shortest path problem makes use of our definition of a string section having constant length independent of its shape. This property permits us to deform a route network we obtain from the map while maintaining the lengths of all route sections. The deformations are constrained by topological and metric conditions which intrinsically guarantee the validity of the operations and their result. By pulling apart the start and end nodes in the network we can straighten the connections between these nodes until one connection is taut; this will be a shortest connection, as there is no shorter connection between two points than a straight line. Fig. 4 illustrates the approach through the depiction of a sequence of states of a physical deformable route network whose start and end nodes are pulled apart. The physical network used for the figure was optically derived from a city map and reconstructed on a 3D printer with non-elastic flexible material.

The example shows how we can represent shapes spatially (instead of formally) and how we can apply operations that by definition maintain the relevant length property. In other words, we can avoid approximations and structural incompatibilities that may be introduced through formalization. We also can avoid approximations we might introduce when interpreting the result of a formal problem solving process in terms of real-world entities. Thus we stick to the inherent spatial structure of the problem domain as best as possible and thus maintain structural fidelity.

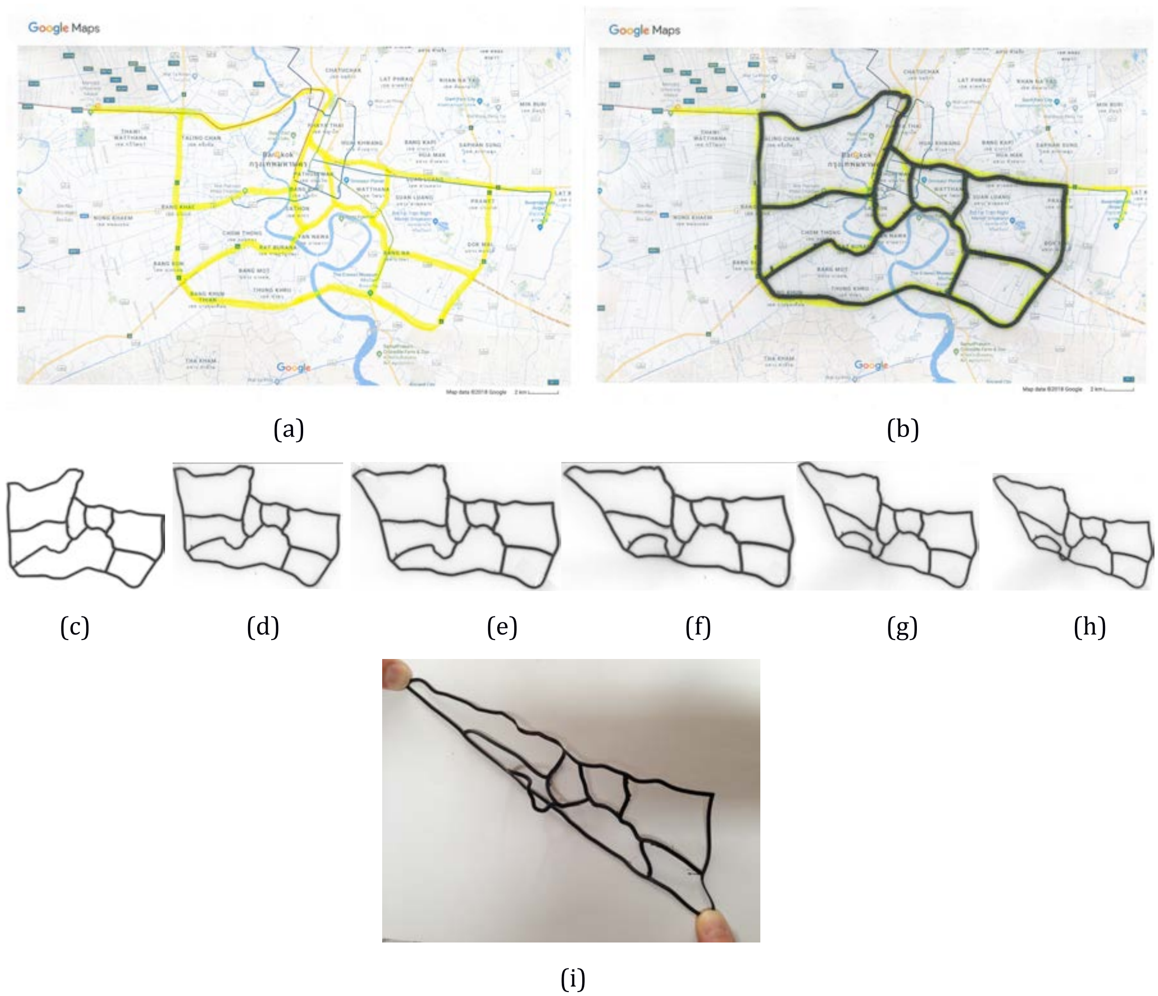

Figure 4. Deforming a route network to identify the shortest path (a) Original route network on the map; (b) flexible non-elastic copy of the network generated by a 3D printer; (c) – (h) sequence of length-maintaining deformation states obtained by pulling apart start and end nodes; and (i) the resulting straight line is the shortest path.

6.3 Trisecting arbitrary angles

Classical compass-and-straightedge geometry permits elaborate constructions involving points, lines, circular arcs, areas, and angles. Viewing a string without pins as a generalized straightedge and a string with pins as a generalized compass, we can conceptualize any geometric construction with strings and pins that we can conceptualize with the compass-and-straightedge metaphor. Strings and pins, however, allow for an even wider range of geometric constructions than compass and straightedge.

Certain geometric constructions cannot be performed with compass and straightedge; specifically it has been shown that it is impossible to trisect arbitrary angles with compass and straightedge alone (Wantzel 1837; Dudley 1987). The Janus-faced nature of strings that serves both as compass and as straightedge enables the strings-and-pins domain to overcome this limitation.

As in classical Euclidean compass-and-straightedge constructions we have to make assumptions regarding the conceptual or physical operations we are allowed to perform. In classical constructions we assume, for example, that the two tips of a compass can be made to coincide with two point locations. Similarly, we will postulate that a string section can be made to coincide with an arc (or other concave shapes). In this way, we will be able to transfer the length of an arc to a string section. (With physical tools such operations are facilitated through the elevation of objects into the third dimension that allow us to snuggle a cord around the objects.)

For trisecting an arbitrary angle, we will exploit the proportionality relation between the central angle α and a corresponding arc as well as the proportionality relation between the radius of a circular sector and its arc length (Fig. 5). We can gauge the arc length s by means of a string to transfer it to an extended circular sector with the same central angle in a circle of triple radius; its corresponding arc will have triple length; by transferring the length of the smaller arc to the larger arc, we get one third of the larger arc; the arc thus obtained defines the central angle α/3.

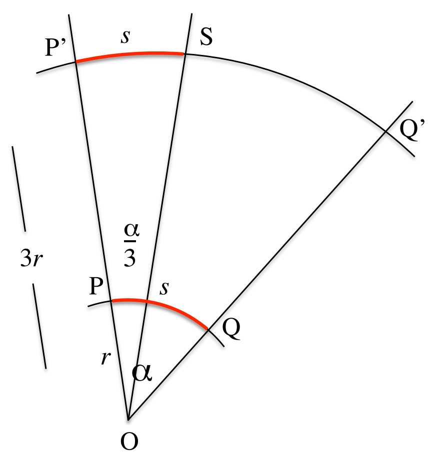

Figure 5. Trisecting angle α by applying length s of arc $\widehat{PQ}$ to a section of arc $\widehat{P'Q'}$

The construction works in detail as follows [Trisecting angle α]:

- construct angle α with arms p and q intersecting at O
- construct arc with center O and radius r
  - the arc intersects p at P and q at Q
  - the arc length $\widehat{PQ}$ equals s
- construct arc with center O and radius $3*r$, arms p and q
  - the arc intersects p at P' and q at Q'
- transfer arc length $\widehat{PQ}$ via string section to arc $\widehat{P'Q'}$ from P' towards Q'
  - the string section ends at S
  - angle P' O S equals α/3

The arc-angle proportionality is well known, of course; it is exploited, for example, in mechanical clocks. Here the angle of the minute hand advances 1/60 of the angle of the second hand and the hour hand advances 1/12 of the angle of the minute hand. In effect, the proportionality of radius and circumference of a circle is used to divide angles. The trisection of arbitrary angles, for example, can be achieved through interlocked cogwheels whose radiuses have a ratio of 1:3. If the small wheel turns by angle α, the large wheel will turn by angle $-\alpha/3$. A compass-like device to trisect angles is presented in Brown (2008).

By the mechanism of gauging the arc of a circle with a string we also can provide a constructive procedure for squaring the circle: we convert the half circle by means of a string to a straight line of length π. We then can apply a standard construction to construct $\sqrt{\pi}$. This provides the side of the square that has the same area as the circle.

7. Features and limitations of constructive and depictive geometry

A main difference between diagrammatic or physical problem solving approaches and abstract (e.g., algebraic) approaches is the *specificity* of the representation. For example, in diagrams, we instantiate specific triangles with edges of specific lengths, with specific angles between the edges, with a specific orientation, etc. This specific instantiation happens in diagrams even if we intend to represent a 'general triangle' with the properties that all triangles share. The specificity is perfectly okay if we have to solve problems that concern specific entities. However, if we intend to make general inferences, we must argue why the special case does not restrict the general validity of the solution (Forbus *et al.* 2011). Polya (1945) advises on getting over the possible "wrong instantiation" fallacy by drawing and trying out a multiplicity of such examples.

From a cognitive perspective, specific and concrete instances make concepts more easily graspable (e.g., Wason 1968; Kahneman 2011); generalization follows later, possibly after exposure to a variety of specific instances (Aamodt and Plaza 1994).

In abstract representations such as algebra, we can let parameters vary in order to represent classes of instances. In fact, the algebraic language is so general that we must take special measures to make sure that what we describe is limited to triangles, by putting explicit constraints on the parameters. In other words, formal representations essentially require that the problem to be solved already is largely understood; otherwise, a correct formalization would be difficult or impossible to achieve.

Thus, in diagrammatic or physical representations we must be concerned about the *generality* of the inferences we may make while in formal approaches we must be concerned about their *specificity*. This illustrates the complementarity of depictive and formal representations, and the potential power of combining them: depictive aspects contribute implicit structures of the target domain and maintain useful domain integrity, whereas formal aspects can transcend these limitations for reasoning about the domain, where the aspects to be reasoned about are individually explicated.

Accordingly, the strengths of the respective approaches will depend on the precise tasks to be solved: do we need the representation to identify and understand the essence of the problem? Do we want to solve a specific spatial problem or are we looking for the solution of a class of problems? How much effort is needed to generalize from a specific spatial solution, or to apply a general abstract solution to a specific spatial situation?

Another aspect to consider is the issue of scaling. Diagrams are used to support thinking and reasoning in the mind of an observer. They extend our rather limited mental imagery abilities, which are restricted by the human working memory capacity, only allowing us to deal with a handful of entities at a time (Miller 1956; Cowan 2001). In diagrams, this limitation is overcome by providing perceptual access to externally represented spatially organized knowledge. But the ranges and bandwidths of perception channels also are limited. Thus, the system consisting of a diagram and an observer is limited by the perceptual field accessible to observers and by their capacity for mental imagery. But if we combine perception abilities with action abilities – such as eye movement, locomotion, and zooming in and out – external spatial media can provide nearly unlimited information about spatial and other feature dimensions. Local and global spatial relations are capable of integrating knowledge of various kinds.

Thus, although mental imagery and spatial perception are focused and limited, there seems to be no principal spatial limit to exploiting 'knowledge in the world' (Norman 1993) for spatial problem solving and reasoning. The integration of a variety of aspects in the structure of the same medium appears to be key to identifying ways for solving spatial problems. Formal representations, on the other hand, serve to analytically disintegrate and linearize the various aspects of knowledge contributing to a problem. They are particularly useful to express and follow up on thoughts and reasoning by an observer and to describe approaches to spatial problem solving once the aspects contributing to the problem have been understood.

In summary: different representations have different strengths. However, these strengths are not yet systematically exploited and employed in a complementary manner. In the domain of spatial problem solving, we propose to analyze not only the kind of knowledge that is to be dealt with, but to consider also (a) the form in which the problem is originally given; (b) the form in which the solution is needed; and (c) the obstacles that need to be overcome in the various phases of the problem solving process, before deciding on suitable tools for the various phases.

8. Epistemological implications and cognitive relevance

In order to identify a larger class of problems that we could investigate regarding their approachability with spatial problem solving methods or metaphors, we looked at the well-established compass-and-straightedge constructions in geometry. We found that any problem we can conceptualize as compass-and-straightedge constructions we also can conceptualize as strings-and-pins construction: the straightedge metaphor is equivalent to the string metaphor when the string is pulled taut and the compass metaphor matches the concept of a taut string section whose one end is fixed by a pin at a given center position and whose other end is marked by a pin at the distance of the radius from the center, that is the boundary line of the circle drawn by a compass. Thus, we use different physical metaphors for identical abstract notions; these metaphors enlarge the scope of operations and interpretations, as they can be subjected to a larger set of physical respectively conceptual operations.

We have shown that we can extend Euclidean constructions in interesting ways that allow us to derive additional spatial relations relevant for everyday problem solving, such as the shortest route problem: Euclidean geometry restricts us to polygonal and circular route shapes. With strings we can directly relate curved shapes to polygonal shapes without going through approximations in formal structures (as when we approximate circular shapes by polygons). In addition, we gain a potential to compare the lengths of arbitrary linear shapes – even shapes that may be difficult to characterize formally. For example, we can map linear boundaries of arbitrary convex objects onto strings by pulling them around these objects; then, we can determine their length by pulling the strings straight and measuring the distance between the ends of the corresponding string section. This approach exploits a spatial equivalence property of differently shaped objects. It is similar to Archimedes' approach to determine the relative purity of his king's gold crown after having an insight regarding the problem when taking a bath in his bathtub (Vitruvius 1914). Here Archimedes exploits the fact that the volume of solid matter is independent of its shape.

In this kind of conceptualization we also can evade a type of problem known as the coastline paradox (Mandelbrot 1983). The coastline paradox demonstrates that the measured length of an unsmooth line depends on the granularity of the measuring device and can increase without limit if measuring resolution gets increasingly finer. It also demonstrates that the formalization of real-world entities may introduce structural incompatibilities in the conceptualization of spatial entities (here due to the discretization of curved lines). Like constructive geometry, the strings-and-pins-based conceptualization avoids discretization; the precision of argument is not limited by the granularity of a representation.

Whereas Euclidean geometry (just like other formal approaches) is not concerned with the question of how to get spatial relations from the real world into formal systems, we propose ways in which we can extend the strictly abstract approach of Euclidean geometry in this direction. We achieve this by applying the notion of a concept not only to abstract entities but also to real-world entities such as roads. The goal of this effort is to identify spatial operations that directly generate solutions within the spatial substrate that otherwise would require abstract representation and computation.

So far, the task of relating a representation to real-world problems is left to human interpreters. We investigate whether we can extend this authority to a larger set of cognitive agents including artificial agents with perception and action capabilities.

The cognitive relevance of our approach is in the need to include *knowledge in the world* as an integral part of cognitive systems, just as natural cognitive agents do it. We postulate that commonsense problem solving in AI will not reach a breakthrough unless they take into account knowledge in the world factually and structurally. We envision a future where the cost of physical robot operations in the world will be very low; then robots will be able to explore space (as we humans have done since childhood) and solve everyday spatial problems. Seeing the knowledge in the world and acting directly on real-world affordances, rather than formalizing everything to compute formal results and pursue the intellectual reasoning approach, can provide shortcuts that could be decisive for the tractability of a problem.

9. Implications on computational spatial problem solving and outlook

We have shown that certain spatial problems can be solved quite directly and efficiently in a spatial medium and we have suggested that we can employ robotic approaches in spatial media to exploit the structural advantages space provides for solving spatial problems. Natural cognitive agents (humans and other animals) who strongly interact with spatial environments by perceiving spatial structures, moving in space, and manipulating spatial configurations heavily employ this kind of approach. Does this mean that we always require physical space to make use of spatial problem solving approaches and that we cannot implement them entirely within computational structures? After all: humans seem to be able to perform certain spatial operations like mental rotation without interaction with their environment.

Our answer is that current computational systems solve spatial problems by decomposing them into linear descriptions of individual spatial relations. This permits the use of list and tree structures that can be processed sequentially. Although all information about spatial relations can be preserved by decomposition and interactions among spatial relations can be simulated through explicit computation, the implicit interaction among spatial relations cannot be realized on the description level of classical computation. It is this implicit interaction that makes our shortest route approach, the use of diagrams and maps, and other spatial operations so efficient, as it offloads computational effort to the spatial substrate.

For a computational implementation of interacting spatial relations we require processing structures that maintain spatial coherence across the two or three spatial dimensions. In other words, when one spatial relation is modified, spatially affected relations need to be modified simultaneously rather than iteratively in order to avoid the computational overhead introduced through sequencing inherently interactive processes. Unfortunately, current approaches to parallel processing are not sufficient, as these require independent activities, whereas spatial processing involves highly dependent processes, and these are responsible for the efficiency. It has been suggested to us that spatial coherence, which our problem solving processes benefit from, may correspond to interactions in quantum computing. We will investigate this connection further.

Sequential processing as in language processing and in reasoning sometimes is associated with the left brain hemisphere, whereas spatial and holistic processing is associated with the right brain hemisphere. We envision the construction of computational processing structures permitting integrated or holistic processing of spatial relations as, for example, in mental rotation. We should add that humans' mental rotation capabilities are rather limited, permitting mental manipulation of only small spatial configurations. Nevertheless, holistic mental manipulation of small configurations may make a big difference to manipulating individual relations sequentially.

In our opinion, a realization of 2D or 3D connectivity structures for small configurations in silicon may be feasible. In this way, we could enable direct interaction of spatial relations. To this end, our approach has been to first investigate different types of spatial approaches in truly spatial substrates. This should help us to better understand which abstract structures and processes will be needed to emulate spatial structures for computational spatial problem solving while preserving the interaction characteristics inherent in spatial substrates.

Acknowledgments

This paper is based on a contribution to *Advances in Cognitive Systems 2016* (Freksa *et al.* 2016). We acknowledge engaged discussions with Pat Langley, Ashok Goel, Ulrich Furbach, Florian Furbach, Ana-Maria Olteţeanu, Holger Schultheis, Frank Dylla, Ahmed Loai Ali, Maria Vasardani, Gabriel Zachmann, and Peter Haddawy. Aaron Sloman and Clayton Lewis pointed us to earlier publications on the shortest route problem than the one we had found. Alexander Mittelstorb transformed parts of the Bangkok city map into a deformable route network. We thank the reviewers of earlier versions of the paper for highly valuable comments and suggestions. We acknowledge support by the German Research Foundation (DFG) under SFB 1320:EASE Project P3 and by the University of Bremen under project CogQDA.

References

Aamodt, A. & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7* (1), 39-59.

Bernays P. (1976). Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. In *Abhandlungen zur Philosophie der Mathematik*, 17-61. Darmstadt: Wissenschaftliche Buchgesellschaft.

Brown, J.R. (2008). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures*. London: Routledge.

Cabalar, P. & Santos, P.E. (2011). Formalising the Fisherman's Folly puzzle. *Artificial Intelligence*, *175*(1), 346-377.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24 (1), 87-114.

Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. Numerische Mathematik. **1**: 269-271.

Dreyfus, H. & Haugeland, J. (1974). The computer as a mistaken model of the mind. In *Philosophy of Psychology,* 247-258. Palgrave Macmillan UK.

Dudley, U. (1987). *A Budget of Trisections*, Springer-Verlag, New York.

Euclid (1926). *The Thirteen Books of Euclid's Elements*, transl. by Heath, T.L. (three volumes) Cambridge University Press.

Finke, R. A. (1989). *Principles of mental imagery*. MIT Press, Cambridge, MA.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, *3* (4), 648-666.

Freksa, C. (1997). Spatial and temporal structures in cognitive processes. In *Foundations of computer science. Potential – theory – cognition,* Springer Berlin Heidelberg, 379-387.

Freksa, C. (2015). Strong spatial cognition. In Fabrikant SI, Raubal M, Bertolotto M, Davies C, Bell S, eds., *Spatial Information Theory* (COSIT 2015), LNCS 9368, 65-86 Heidelberg: Springer.

Freksa, C., Barkowsky, T., Dylla, F., Falomir, Z., Olteteanu, A.-M., van de Ven, J. (2018). Spatial problem solving and cognition, in: Zacks, J., Taylor, H., eds., *Representations in Mind and World,* 156-183, New York: Routledge.

Freksa, C., Olteteanu, A.-M., Ali, A.L., Barkowsky, T., van de Ven, J., Dylla, F., Falomir, Z. (2016). Towards spatial reasoning with strings and pins, *Advances in Cognitive Systems* 4, Poster Collection #22, 1-15, http://www.cogsys.org/papers/ACS2016/Posters/Freksa_et.al-ACS-2016.pdf

Giaquinto, M. (2007). *Visual thinking in mathematics: An epistemological study*. Oxford: Oxford University Press.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kozhevnikov, M., Motes, M.A. & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31* (4), 549-579.

Kuipers, B. (1994). *Qualitative reasoning: Modeling and simulation with incomplete knowledge*. The MIT Press.

Larkin, J.H. & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65-100.

Mandelbrot, B.B. (1983). How long is the coast of Britain. Ch. 5 in *The Fractal Geometry of Nature.* New York: W. H. Freeman, 25-33.

Manders, K. (2008). The Euclidean diagram. In Mancosu, P., ed., *Philosophy of Mathematical Practice*, Oxford: Clarendon Press, 112-183.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing, *Psychological Review*, 63 (2), 81-97.

Minty, G.J. (1957). A comment on the shortest-route problem, Operations Research 5, 724.

Mumma, J. (2012). Constructive geometrical reasoning and diagrams. *Synthese*, *186* (1), 103-119.

Norman, D.A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action. *Cognitive Science,* 17 (1), 1-6.

Palmer, S.E. (1978). Fundamental aspects of cognitive representation. In: Rosch, E., Lloyd, B.B. (eds.) *Cognition and categorization*, 259-303. Lawrence Erlbaum, Hillsdale

Polya, G. (1945). *How to solve it*, Princeton: Princeton University Press.

Scaife, M. & Rogers, Y. (1996). External cognition: how do graphical representations work? *Int. J. Human – Computer Studies* **45** , 185-213.

Sloman, A. (1985). Why we need many knowledge representation formalisms. In: Bramer,M, ed, *Research and Development in Expert Systems*, 163-183. New York: Cambridge University Press.

Vitruvius, M. (1914). *The Ten Books on Architecture*. Translation by Morris Hicky Morgan, Harvard University Press, Cambridge, 253-254.

Wantzel, P.M.L. (1837). Recherches sur les moyens de reconnaître si un problème de Géométrie peut se résoudre avec la règle et le compas. *Journal de Mathématiques Pures et Appliquées*. 1. **2**: 366–372.

Wason, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20 (3): 273-281.