Scalable Representation Structures for Visuo-Spatial Reasoning — Dynamic Explorations into Knowledge Types

Sven Bertel  
Human Factors Division & Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, USA

Jan Frederik Sima and Maren Lindner  
Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, Universit"at Bremen, Germany

Abstract

A sizable fraction of current research into human visuo-spatial knowledge processing explicitly or implicitly suggests a spatial processing of certain knowledge types and a visual processing of others. Similarly, many formal and technical approaches for representing and processing visuo-spatial information in artificial intelligence, in computational cognitive modeling, or in knowledge representation and reasoning explicitly or implicitly treat visual and spatial information as belonging to separate types. While there exists good evidence for some differences in mental processing of different visuo-spatial knowledge types, there is much less reason to maintain the currently ascribed separation between the visual and the spatial.

We provide arguments on why strict dichotomies seem unwarranted with regard to descriptions of human mental spatial reasoning and disadvantageous for the formal and technical approaches. We build upon a synopsis of psychological evidence for the existence of multiple knowledge type specific representations in human visuo-spatial reasoning and discuss the notion of scalable representation structures. In absence of proof to the contrary, it seems better practice to assume that (a) many of the type differences attributed to visuo-spatial knowledge processing are gradual rather than qualitative in nature, and that (b) tasks involving visuo-spatial knowledge of several types are often mentally processed through dynamic associations of structures for processing basal knowledge types.

The paper calls for more investigations of human reasoning in visuo-spatial tasks in which knowledge types dynamically change during reasoning. It outlines a research framework for systematically investigating different basal visuo-spatial knowledge types and their combinations with regard to cognitive and computational plausibility. Current research is related to the framework, including research on Casimir, our computational cognitive architecture for reasoning with visuo-spatial knowledge. We argue that a more systematic course of research along the lines of the proposed framework will not only lead to more appropriate descriptions of human cognition (regarding visuo-spatial knowledge processing) but may also spawn more integrated and versatile formal and technical approaches for dealing with visuo-spatial information.

The Visual and the Spatial

In this article, we will address a few issues that concern our current understanding of knowledge types frequently described to be involved in mental reasoning with visuo-spatial knowledge. In particular, we will focus on the interrelations between knowledge types, such as between different spatial types (e.g., holding knowledge about directions, distances, topology) as well as between these and types often described as visual. Second, we will suggest a novel research framework which we believe to be quite capable of bringing about a much more accurate understanding of which knowledge types are actually involved in mental visuo-spatial reasoning and under which conditions, and of how visuo-spatial knowledge gets dynamically maintained and integrated within and across types during reasoning.

Various past and ongoing research addressed and addresses questions related to sensory modality in reasoning with visuo-spatial knowledge. For example, are the involved mental representation structures more adequately described as visual, spatial, motor-related, etc., as propositional, or as abstract and non-modal? What is the involvement of visual or motor imagery processes in mental reasoning with visuo-spatial knowledge, and which subsystems of working memory are drawn upon under which circumstances?

A rough grouping of existing research identifies two main camps: one includes modalist accounts which describe the involved mental representations, at least partially, in terms of mental images (Bertel et al. 2006; Kosslyn 1994), through analogies drawn from diagrams (Chandrasekaran et al. 2004), and as tapping into and relying on mental visuo-spatial subsystems (Kosslyn and Thompson 2003; Ishai and Sagi 1997). This group further involves motor-based accounts that, for instance, explain aspects of visuo-spatial reasoning through enacted or reenacted shifts of eye fixations or visual attention (Richardson and Spivey 2000; Laeng and Teodorescu 2002; Johansson, Holsanova, and Holmqvist 2005) or draw connections between success in reasoning and an embodying of the solution in terms of movements or attention shifts (Thomas and Lleras 2009).

The second group includes non-modal accounts which are, among others, rooted in analogies to propositions (Jahn...
2003) or formal spatial structures such as lists or arrays (Ragni, Knauff, and Nebel 2005), often draw on descriptions of reasoning in terms of spatial mental models (Johnson-Laird 1983) and mainly of central executive functions of working memory rather than of those of modal subsystems (Gilhooly et al. 1993).

Let us start with a quick inspection of evidence for differences in processing: Neuropsychological findings support a separation in higher-level visual processing along two major pathways: a ventrally projecting, occipital-to-inferior-temporal pathway that processes object properties such as shape or color, and a dorsal, occipital-to-posterior-parietal pathway for processing spatial attributes and movements (Ungerleider and Mishkin 1982; Haxby et al. 1991). Furthermore, double dissociations seem to exist between visual and spatial short-term memory (Klauer and Zhao 2004) and research on inter-individual differences in use of how mental representations suggests different individual cognitive styles with preferences for visual, spatial, verbal or other representations (Hegarty and Kozhevnikov 1999). Within much of the research on mental visuo-spatial reasoning, the use of visual (often modal) and spatial (often amodal) representations and reasoning strategies has often been either contrasted (Knauff and Johnson-Laird 2002), or only one representation type has been studied instead of both.

It has been previously argued (Schultheis et al. 2007) that, at least for spatial knowledge processing, the predominant focus on dissociating visual and spatial knowledge processing has led to several misconceptions about the involved mental representations and that, again at least for spatial knowledge processing, the visual and the spatial should better be conceptualized as falling within a continuous dimension of representation. A specific mental representation does thus possess visual as well as spatial aspects. In a nutshell, the argument presented by Schultheis et al. is twofold: First, there exists no consensus in current psychological research as to what constitutes a spatial and a visual mental representation, as to which entities are processed spatially or visually, and as to which traits indicate an individual’s inclination to process information visually or spatially. Second, the strict separation between spatial and visual representations (or, for that matter, a concurrent use of separate representations from both types) found in current artificial intelligence systems and computational models of mental spatial knowledge processing is computationally inefficient and cognitively implausible.

Schultheis and colleagues suggest four criteria for determining where inside the visuo-spatial continuum a specific representation or represented object falls. A representation or an object is the closer to the visual extreme of the continuum the more types of spatial relations are involved (i.e., the more different spatial knowledge types such as distance, topology or orientation have to be integrated, the more visual the representation gets), the more relations are involved, the more specific (i.e., less ambiguous) the representation is, and the higher the exemplarity of its content (such as for shapes, color, texture, etc.) and the lesser prototypical, categorical information it contains. Fig. 1 shows an example of a pair of representations that both represent the same spatial situation. According to all four criteria, Fig. 1b rates as being closer to the visual end of the continuum: it represents more spatial relations than Fig. 1a (e.g., the fact that the pond is in front of the tree) and more types of spatial relations (e.g., also distance information), it is more specific (e.g., regarding the precise positions of the objects) and exhibits a higher degree of exemplarity for the objects.

**Scalable Visuo-Spatial Representation Structures**

If spatial knowledge processing which involves representations that are neither exclusively spatial nor visual but combine spatial and visual traits to different degrees has in fact advantages in terms of computational and plausibility properties, we should certainly try to find out more about the characteristics of the involved processes and representations. Ideally, such representational systems (i.e., the set of representation structures and associated processes) should permit an incremental construction of increasingly visual structures by starting off spatial (i.e., low-visual) ones and adding more and more spatial relations, relation types etc. The exact visuo-spatial characteristics of the representational system are thus dynamically determined; knowledge types etc. can be added to the representation on demand. Schultheis et al. suggested to use the term **scalable representation structures**. Current computational models of human visuo-spatial reasoning as well as formal and technical approaches to processing visuo-spatial information also make use of representation structures that differ in number and kinds of employed visuo-spatial knowledge types. That is to say, they are either used to represent one specific type of knowledge or a specific fixed combination of knowledge types. The distinctive property we propose is scalability in the sense that knowledge types can be added to the representation on demand. This is something not offered by current models. A scalable representation structure as described above can be imagined in two ways:

1. as one single structure that is able to (potentially) accommodate all types and numbers of spatial representations at once, or
2. as a composite structure that consists of a set of distinct, basal, knowledge type specific representation structures which get dynamically linked together in associations as demanded by the task.

It turns out that a synopsis of existing psychological as well as artificial intelligence and cognitive science research is instrumental in determining the more plausible of the two options. Psychological evidence seems to indicate the exis-
tence of not just one single representation structure that is involved in the different aspects of human spatial cognition, such as would be implicated by the first option. Rather, it supports the assumption that spatial knowledge about the environment is stored in the form of several “...smaller chunks, each of which is encoded by a separate representation” (Brockmole and Wang 2002). As we cannot present a detailed survey here for reasons of limited space, we will name but a few selected findings. As Schultheis et al. point out, the existence of different mental representations has been suggested with respect to (a) the representational content (i.e., different hierarchically representations; see, e.g., (Hirtle and Jonides 1985); (McNamara 1986)), (b) whether object information (what) or locational information (where) is represented (Hegarty and Kozhevnikov 1999; Klauer and Zhao 2004), and (c) whether relations are represented egocentrically or allocentrically (Easton and Sholl 1995; Rieser 1989). Corroborating evidence from imaging studies includes, among other findings, evidence for specific retinotopic structures beyond occipital cortical regions, such as in superior parietal cortex for holding spatial knowledge on angles from previously seen contralateral visual targets (Sereno, Pitazlis, and Martinez 2001).

Cognitive science and AI literature further supports the second option, as numerous different types of spatial representation structures have been proposed, for example, for diagrammatic reasoning (Chandrasekaran et al. 2004) and computational cognitive modeling of spatial reasoning (Ragni, Knauff, and Nebel 2005). From a representation theory point of view, it has been claimed that having a set of several special-purpose representation structures has advantages over having just a single, general-purpose one (Sloman 1985). Interestingly, no technical or formal system or cognitive model exists yet that employs just a single, encompassing representation structure in the sense of the description of the first option.

Integration of Multiple Knowledge Types

In order to formalize the concept and implement a computational model of the kind of scalable representations structure along the lines described above, we need to turn to matters of dynamic integration of different knowledge types held in a distributed fashion across several knowledge type-specific representation structures. Any design of a suitable set of integration or association processes needs to consider requirements for arbitrary combinations of different types of knowledge, or for different amounts of knowledge for that matter, (e.g., for keeping global or local consistency) as well as consider implications of the concrete combination cases (such when integrating knowledge about topology with knowledge about orientation). Astonishingly few approaches exist that specifically target issues of integrating type-specific representation structures. Existing approaches and models of spatial reasoning from fields such as cognitive modeling (Ragni, Knauff, and Nebel 2005; Goodwin and Johnson-Laird 2005), Geographical Information Systems (Sharma 1996) and formal reasoning (Egenhofer 1994; Frank 1996; Moratz, Nebel, and Freksa 2003) still all share the problem of being structurally restricted to reasoning with either one specific kind of spatial knowledge or the combination of just two specific knowledge types.

A first step towards implementing our concept of a scalable representation structure for visuo-spatial knowledge processing consisted in identifying dependencies between knowledge of different types (Schultheis et al. 2007). We will look at one example here: Let us assume we know that some object X is a proper part of another object Y, which is a topological knowledge fact, and that we further know that Y is far from some third object Z, which is distance knowledge. In such case, the distance relation between X and Z will not be fully unspecified; rather, X has to be far from Z. We identified various dependencies such as this one across different types of spatial knowledge and created and analyzed appropriate composition tables (see Table 1, for an example). These dependencies likely need to be taken into account whenever a new type of knowledge or just a new fact is included in an existing visuo-spatial mental representation. The knowledge represented in the structure then needs to be adjusted accordingly. We assume that the underlying mechanism is best realized by mutual processes for each pair of knowledge types currently employed in the representation structure.

Regarding matters of cognitive adequacy of our concept, we need to find answers to several, yet insufficiently answered questions. Two of the more pressing ones are: (1) Which combinations of visuo-spatial knowledge are cognitively plausible (i.e., are actually used by humans)? Likely not all possible combinations will be found in human visuo-spatial reasoning. (2) To which extent is knowledge kept consistent during human visuo-spatial reasoning when integrating knowledge of different knowledge types? How can we systematically use findings on inconsistencies of mental knowledge processing to find out more about the underlying integration / association processes?

Casimir: A Cognitive Architecture for Spatial Reasoning

Our concept of scalable representations structures for visuo-spatial reasoning is part of the development of our cognitive architecture Casimir – Computational Architecture, Specification and Implementation of Mental Image-based Reasoning (Barkowsky 2007). Casimir aims at providing a cognitively plausible computational description for the processes and representations involved in mental reasoning with mental images and mental models. As a cognitive architecture, Casimir is designed as a comprehensive system which includes mental processes in long-term memory, processes in working memory as well as models of the interaction between mental representations and external diagrams. Casimir therefore provides an architectural outline for the integration of different modeling aspects rather than as a specific computational model. Compared to other cognitive architectures, Casimir possesses a clear emphasis on the processing of visuo-spatial knowledge and on the development of adequate working memory structures for mental spatio-
analogue representations. It further provides the possibility to externalize the content of these analogue representations into diagrammatic media as well as of using these media as input for internal reasoning. Such combination of features have currently not been implemented in any other cognitive architecture.

**Need for Dynamic Explorations into Knowledge Types**

Let us recapitulate before we will make the case for a new line research into visual-spatial mental knowledge processing: There exists evidence for differences in mental representation and processing for a number of basal spatial knowledge types. Also, some visual-spatial reasoning tasks seem to involve cognitive subsystems generally associated with visual mental imagery while other tasks do not.

One suitable research strategy for shedding light onto the involved mental representations involves testing within classes of visual-spatial tasks that one assumes to tap into the same (or similar) cognitive resources and that involve the same or similar types of spatial knowledge. We can in fact find various instantiations of this strategy: For example, over the last years, many aspects of deductive reasoning about directional knowledge between physical objects in table-top scenarios have been thoroughly investigated; the result have been described within mental model-based theories and have been related to imagery (Knauff and Johnson-Laird 2002). Eventually, when ones combines the results of different research with this strategy, one may produce a table of mappings between types of visual-spatial tasks on the one hand and involved aspects of mental representation and processing on the other. The dimensions on which the tasks can be varied include the set of involved relation and object types as well as the cardinality of involved relations and objects. Consequently, a cognitive model may be produced for each class of types that share common representational and procedural properties.

This research strategy has a number of advantages as it, for instance, permits relatively homogeneous experimental paradigms within the individual types of visuo-spatial tasks, thereby greatly facilitating comparisons between experimental conditions as well as theory and model building. The downside of this strategy is, however, that, first, many different theories and models are produced whose explanatory powers are each limited to a specific class of tasks. Second, as most authors search for and seem to like to identify general principles of mental knowledge processing and representations, many results from small-scope experimental settings often are overly generalized, leading to a host of theories and models about human reasoning with visuo-spatial knowledge, many of which can be expected to be conflicting, inaccurate and hard to compare to other theories and models because of the specificity of the class of tasks from which they were derived. The authors of the current paper are no exception and similarly enjoy generalizing, so we contend that this problem in fact holds for many areas of cognition research, not just for visuo-spatial reasoning.

Last, this research strategy de-emphasizes research on tasks in which the dimensions that were mentioned above are not kept stable over the course of the reasoning but vary instead. Imagine, for example, a task in which subjects are first asked to integrate a set of directional spatial relations between objects which are subsequently supplemented by information about distances between some of the objects, or about topology. Does this then lead to a recoding of mental representations in terms of different mental representation formats (with associated different processes) and are the existing mental representations simply joined and extended by additional ones? It is our impression that as most of the existing research follows the strategy described above, questions such as this one are hard to answer. We argue that an additional research strategy is needed that focuses on the dynamics of the knowledge type structure, in particular, on cross-type growth and shrinkage of a currently held visual-spatial knowledge base.

We suggest to start with compiling a list of basal spatial knowledge types for whose involvement (and representation) in human reasoning good evidence can be found in cognition research. Examples may include knowledge about cardinal directions, angular information, distances, various aspects of topology, as well as form or color knowledge. This list should be checked for computational plausibility against findings from AI, knowledge representation and reasoning research. Should the check suggest that additional spatial basal knowledge types may be involved in human visuo-spatial reasoning for which no or no conclusive psychological evidence yet exists, then new cognition research paradigms need to be devised and implemented to help clarify the possible existence of basal representation structures for these additional knowledge types.

Second, we should look for evidence for the existence of associations of basal knowledge type pairs (later: triplets, etc.), such as between knowledge about cardinal direction and distance knowledge. Which classes of tasks require

<table>
<thead>
<tr>
<th>$X \cap \exists_{Z} Y \cap \exists_{X}$</th>
<th>close</th>
<th>medium</th>
<th>far</th>
</tr>
</thead>
<tbody>
<tr>
<td>equal</td>
<td>$TOP$</td>
<td>disjoint</td>
<td>disjoint</td>
</tr>
<tr>
<td>disjoint</td>
<td>$TOP$</td>
<td>$TOP$</td>
<td>$TOP$</td>
</tr>
<tr>
<td>tangent</td>
<td>$TOP$</td>
<td>${\text{equal, in t/nt}}$</td>
<td>${\text{equal, in t/nt}}$</td>
</tr>
<tr>
<td>overlaps</td>
<td>${\text{equal, in t/nt}}$</td>
<td>${\text{equal, in t/nt}}$</td>
<td>${\text{equal, in t/nt}}$</td>
</tr>
<tr>
<td>in/t</td>
<td>$TOP$</td>
<td>disjoint</td>
<td>disjoint</td>
</tr>
<tr>
<td>in/nt</td>
<td>$TOP$</td>
<td>disjoint</td>
<td>disjoint</td>
</tr>
<tr>
<td>contains/t</td>
<td>$TOP$</td>
<td>${\text{equal, in t/nt}}$</td>
<td>${\text{equal, in t/nt}}$</td>
</tr>
<tr>
<td>contains/nt</td>
<td>$TOP$</td>
<td>${\text{equal, in t/nt}}$</td>
<td>${\text{equal, in t/nt}}$</td>
</tr>
</tbody>
</table>

Table 1: Composition table showing the possible topological relations between $Y$ and $Z$ (shown in the corresponding table cells) given both a topological relation between $Y$ and $X$ ($Y \cap \exists_{X}$) and a distance relation between $X$ and $Z$ ($X \cap \exists_{Z}$); $TOP$ is the set of all possible topological relations, as defined by the RCC-8 calculus (Randell, Cui, and Cohn 1992).
which combinations of knowledge classes? We should best try to aim to obtain findings for closely neighboring sets of knowledge types within the suggested visuo-spatial representational continuum. So far, many studies that involved tasks requiring mental representations from different parts of the continuum chose representation pairs that are located on opposite ends of the continuum in order to show clear dissociations. Following our line of argument from above requires putting at least as much attention on similarities of different types of representation than has been put so far on their differences.

Most importantly, however, we will need to adapt our research strategies to include visuo-spatial tasks that require varying the involved relation types, number of objects, number of relations, and object types during the reasoning. It will be only with such tasks that we will be able to clarify the existence and characteristics of the postulated scalable mental representation structure. Specifically, we hope to be able to observe whether the constructed and maintained mental representations grow continuously or whether their growth exhibits discontinuities (for instance, as shown by step functions in measured reaction times, error rates, etc.). Should discontinuities be found, they may be indicative of mental recoding processes that may be triggered once a certain threshold of the number of involved relations or relation types is reached. It is only with tasks that dynamically change representational requirements that we can adequately investigate the interrelations and transitions between mental representations of different knowledge classes and eventually increase our understanding of the basic cognitive dynamics of associating heterogenous knowledge types.

We argue that the potential implications of the proposed research framework go far beyond a chance for greatly improving our knowledge about the dynamic functioning of mental visuo-spatial reasoning. A better understanding of how humans can successfully cope with dynamically changing spatial situations while keeping largely stable representations has great potential to inspire technical and formal approaches to visuo-spatial information processing. We should keep in mind that many of the current formal approaches to spatial knowledge representation and reasoning are particularly lacking functionality in the integration of information across types.

Conclusions

This paper makes the case for a new research framework that specifically focusses on dynamic aspects of visuo-spatial knowledge representation and reasoning. We provided selected psychological evidence that human visuo-spatial reasoning is based on different task-dependent mental representations, but that these should not be seen and modeled as distinct and independent representations, as it is the case in current cognitive, computational or formal approaches (most prominently, dividing spatial and visual representations). Instead, we propose the concept of a scalable representation structure that allows for a dynamic combination of basal, knowledge type-specific representations. We outlined the necessary mechanisms and processes that need to be further investigated on a computational and cognitive level to fully examine the cognitive plausibility of such a structure and to seek to implement reasoning with it for computational cognitive modeling as well as for technical application.

Acknowledgement

This contribution includes the presentation of work done in the R1-[ImageSpace] project of the Transregional Collaborative Center SFB/TR 8 Spatial Cognition. Funding by the German Research Foundation (DFG) is gratefully acknowledged.

References


