# Stop Using Introspection to Gather Data for the Design of Computational Modeling and Spatial Assistance

## Markus Knauff

Max-Planck-Institute for Biological Cybernetics
Center for Cognitive Science, University of Freiburg
Transregional Collaborative Research Center on Spatial Cognition (SFB/TR 8)
*markus.knauff@tuebingen.mpg.de*

### Abstract

Research in spatial cognition offers a wide spectrum of possibilities to combine psychological and computational approaches. Sometimes, the design of computational systems may merely be based on what researchers believe happen in their minds when solving spatial reasoning problems. In other cases, an already existing computational approach is enabled post hoc as cognitively adequate, or psychologically valid. In the contribution, it will be argued that carefully conducted experiments with human participants are needed to understand spatial cognition. Based on such experiments, the contribution will then clarify the role of visual images and spatial representations in human reasoning and problem-solving. What is represented in such representations? Visual features such as color, shape, texture, etc? Spatial information such as topological, metrical, or ordinal relations? Then the contribution will provide experimental evidence on which kind of information should be presented in diagrams so that they efficiently support reasoning processes in humans and technical systems, as well as in human-machine interaction. Finally, it will provide some thought on why reasoning could be so inextricably connected to the cognitive representation and processing of space.

## If you ask people how they reason …

… many of them say that they rely on visual mental images. They often say that they form a mental picture in their ''mind's eye'' and then bok at this picture to find new information. Not only non-psychologists, but also many cognitive psychologists have claimed that reasoning is inextricably linked to seeing in the ''mind's eye'' (e.g., De Soto, London, & Handel, 1965; Kosslyn, 1994). So, do diagrams similar to visual images support reasoning in humans, technical systems, and in human-machine interaction? The aim of the contribution will be to discuss the pros and cons of this view. For this purpose, the hypothesis that visual representations underlie reasoning will be re-examined and discussed from a cognitive and a neuro-cognitive perspective. Then the contribution will focus on spatial representations as an alternative account to explain human thinking. Are individuals able to use "spatial imagery" in the sense that they represent visual features only if they are necessary for the task, but otherwise focus on spatial information only? If so, what does that mean for the support of reasoning processes in humans by technical systems? The contribution will report on experimental findings from different laboratories including the research from the contributor's own lab. Then it will outline some consequences with respect of the issue of the AAAI Workshop. Finally, it will provide some thought on why reasoning could be so inextricably connected to the cognitive representation and processing of space. The following briefly summarizes the line of argumentation that will be contributed to the workshop.

## Stop using "introspection" to gather data for the design of AI-systems

Herbert Simon and Allen Newell identified the 'thinking-aloud protocol" as a method used to gather data in the design and development of AI-systems and software engineering. Basically, this involves simply asking humans to think aloud as they are performing a set of specified tasks. The investigator views the individual who speaks his thoughts aloud, thus allowing study of his introspection. Although there is no doubt that introspection, the observation of one's own inner life, can be a fruitful source of initial inspiration, it cannot be expressed often enough that it can be (and as a matter of fact frequently is) fatally misguiding. One of many disadvantages is that people typically do not distinguish between different types of introspections: representational states and cognitive operations. The distinction between the two representational states (i.e., the mental representation of entities or events in their absence) and cognitive operations (i.e., construction, rehearsal, inspection, and manipulation of such representations) are absolutely essential if we try to design computational systems that work in a human-like fashion. The only way to gather such information is – at least from the perspective of an experimental psychologist – to conduct carefully designed

experiments with human participants. Such experiments can use performance measures or neural activities as manifestation of the underlying mental representations and cognitive processes.

## Psychological "facts", Part 1: The visual-impedance effect

Although it is natural to suppose that visual mental imagery is important in human reasoning (and thus diagrams consisting of visual features can support human performance) the evidence is equivocal. However, in Knauff & Johnson-Laird (2002), it was argued that researchers often do not distinguish between ease of visualization and ease of constructing spatial representations. Rating studies, however, show that these factors can be separated. Their results yielded four sorts of relations: (1) visuo-spatial relations that are easy to envisage visually and spatially, (2) visual relations that are easy to envisage visually but hard to envisage spatially, (3) spatial relations that are hard to envisage visually but easy to envisage spatially, and (4) control relations that are hard to envisage both visually and spatially. Three experiments by Knauff & Johnson-Laird (2002) showed that visual relations slow down the process of reasoning in comparison with control relations, whereas visuo-spatial and spatial relations yield inferences comparable with those of control relations. This speaks against the "orthodox" visual imagery theory and shows that irrelevant visual detail can even be a nuisance in reasoning and can impede the process. This also has consequences for the development of spatial assistance systems: Do not overestimate the role of visual information!

## Psychological "facts", Part 2: The spatial-interference effect

Other data also support the role of spatial representations and processes, but question the role of visual images. In Knauff, Strube, Jola, Rauh, and Schlieder (2004), we explored thinking with spatial relations and whether the underlying processes are inherently visual, spatial, or logical. We applied the dual task interference paradigm to spatial reasoning problems in one dimension, using Allen's interval calculus, in order to make progress towards resolving this argument. We showed that previous reasoning experiments in the dual-task paradigm confounded visual and spatial components in the secondary tasks. We then reported two experiments, in which participants solved relational reasoning problems alone and while performing one of four secondary tasks: 1. visual secondary tasks which were visual but not spatial, 2. visuo-spatial secondary tasks which were visual and spatial, 3. spatial secondary tasks which were spatial but not visual, and 4. control secondary tasks which were neither visual nor spatial. The experiments show that visuo-spatial and spatial secondary tasks interfere with reasoning, whereas reasoning performance together with visual and control secondary tasks were comparable to undisrupted reasoning. These findings indicate that reasoning and spatial secondary tasks utilize the same subsystem of working memory, whereas visual subsystems don't appear to be involved. Our results have important consequences for the development of spatial assistance systems. They show that reasoning can be impeded if a system provides information that interferes with the reasoning process itself.

## Psychological "facts", Part 3: Results from experiments with sighted, blindfolded, and congenitally totally blind people

The visual nature of reasoning might suggest that congenitally totally blind individuals—that do not experience visual mental images—should be impaired in reasoning with highly visual materials. On the other hand, there are several studies showing that people who are blind from birth differ from sighted people in their use of visual images, but that they are as good as the sighted in the construction of spatial representations (e.g. Kerr, 1983). In Knauff & May (2004, submitted) we again used three sorts of relations: (1) visuo-spatial relations, (2) visual relations, and (3) control relations. In absolute terms, congenitally totally blind people performed less accurately and more slowly than the sighted on all such tasks. In relative terms, however, the visual relations in comparison with control relations impeded the reasoning of sighted and blindfolded participants, whereas congenitally totally blind participants performed the same with the different sorts of relations. We concluded that mental images containing visual details that are irrelevant to an inference can even impede the process of reasoning. Persons who are blind from birth—and thus do not tend to construct visual mental images—are immune to this visual-impedance effect.

## Psychological "facts", Part 4: Preferred mental models

In Knauff, Rauh, and Schlieder, (1995), an experiment based on Allen's calculus (Allen 1983) and its transfer to qualitative spatial reasoning was conducted. Subjects had to find a conclusion X r3 Z that was consistent with the given premises X r1 Y and Y r2 Z. The results support the assumption that there are preferred models when people solve spatial three-term series. Individuals draw just one conclusion, although other logically correct conclusions exist. Based on this finding we conducted a whole series of additional experiments. In Rauh, Hagen, Knauff, Kuß, Schlieder, and Strube (2005) we report that inferences that conform to these preferred mental models (PMM) are verified faster and more often correctly than inferences that are valid for alternatives. We also show that individuals perform a revision process to make sure that no other model exists that refutes the conclusion. This process usually

starts with the PMM and then constructs alternative models by local transformations. Models which are difficult to reach are thus more likely to be neglected than models which are only minor revisions of the PMM. The findings resolve a number of open questions in the psychology of reasoning and show that human reasoning processes are based on spatial representations and processes. Explanations based on visual mental images seem to be unable to account for the data. The data are also of interest for areas of spatial information systems since the interval relations of Allen (1983) are used.

## Psychological "facts", Part 5: Spatial reasoning and the involvement of visual brain areas

In Knauff, Fangmeier, Ruff, and Johnson-Laird (2003), we investigated the neurocognitive processes of mental imagery in reasoning. We again used the above-mentioned relations (visual, visuo-spatial, spatial, and control) and conducted an experiment using functional magnetic resonance imaging. This experiment showed that in the absence of any correlated visual input (problems were presented acoustically via headphones), all types of reasoning problems evoked activity in the right superior parietal cortex, and bilaterally in the precuneus – areas typically related to spatial processes. However, only the problems based on visual relations also activated areas of the visual association cortex corresponding to V2. The results indicate that cortical activity during reasoning depends on the nature of verbal relations. All relations elicit mental representations that underlie reasoning, but visual relations can in addition elicit visual images. Interestingly, these images can even be a nuisance in reasoning and impede the process (see below). These findings agree with the results of Knauff, Mulack, Kassubek, Salih, and Greenlee (2002).

## Psychological "facts", Part 6: Visual brain areas are not essential for reasoning

Fangmeier, Knauff, Ruff, Sloutsky (submitted) conducted an event-related fMRI study, in which we investigated the neuro-cognitive processes of three temporally separable phases of spatial reasoning: (1) understanding the meaning of premises, (2) generating new, not explicitly given information, and (3) deciding whether a conclusion logically follows from the premises. We found three distinct patterns of cortical activity during these phases, with initial temporo-occipital activation shifting to prefrontal and then parietal cortex during the reasoning process. These findings show that most of the brain areas were associated with only one phase of the inference process, thus demonstrating that the sub-tasks tap largely independent components of logical thinking. The most important result is that visual activation was only obtained as a side effect of premise processing. The reasoning process itself relied upon modality-independent spatial representations during the validation of the premises. These areas play a crucial role in spatial processing and working memory and in the integration of sensory information from all modalities into egocentric spatial representations. This account also resolves inconsistencies in previous neuroimaging studies on reasoning. These studies have similarly implied that the parietal cortex may play a key role in reasoning based on mental models, which are supposed to be of abstract spatial nature. However, the results of studies have often shown concurrent activation of visual association cortices. One could argue that this points to the role of "visual mental imagery" in reasoning. In contrast, the study showed that visual association areas are indeed involved in premise processing and the construction of an initial static representation of the initial model, but that more abstract spatial representations held in parietal cortices are crucial for subsequent processes.

## Psychological "facts", Part 7: Individual differences in reasoning

In Ruff, Knauff, Fangmeier, and Spreer (2003), we measured brain activity with fMRI, measured the participants' spatial ability with a well-known subset of tasks from an intelligence inventory, and interviewed the participants on how they solved spatial reasoning problems. All subjects reported having used a more spatial strategy for solving the reasoning problems. They reported having integrated the three objects into a single model, and differences between subjects were only present in the level of detail of the representation (e.g., whether they explicitly contained all three colors or not). Another interesting result was that the brain activation was significantly modulated by the participants' spatial skill. Remarkably, these clusters were located in the precuneus and displayed a negative covariation with the participants' spatial ability score. The higher the visuo-spatial skill, the less activation was measured in that area. No other areas displayed such a relationship. In other words, the higher the participants' spatial skill, the better their reasoning performance, and the less activation present in the precuneus during reasoning. This pattern agrees with recent findings on the effects of skill level on neuronal activity; the reasoning problems seemed to have placed less demand on the spatial processing resources of participants with high skill levels, so that less activity in the relevant cortical regions was required.

## Speculations on why reasoning could be so inextricably connected to space

The reported psychological "facts" highlight the important role of spatial representations and can be interpreted as a warning not to overestimate the role of visual images. But why could reasoning be so inextricably connected to the

cognitive representation and processing of space? What's more, how might humanity have come to develop such spatial strategies to solve reasoning problems in daily life? Overall, the contribution focuses on both aspects of human introspection: representations and operations. It will be argued that the representation and inspection of inherently perceptual visual images typically experienced during thinking might not be as important as it has frequently been suggested. In contrast, many experimental findings from the contributor's own laboratory and from other groups (I apologize that I have cited only just a few in this paper), draw a reasonably consistent picture that seems to be emerging across a number of different experimental paradigms and computational considerations. In this picture, amodal spatial representations and processes turned out to be the vital foundation of human reasoning. At present, the answers to above-mentioned questions can only be tentative. One possible response to the why question might be that reasoning — like fundamental memory processes — has to be performed extremely quickly and accurately, and must be sheltered from external disruptions. Besides, in contrast to the mental rotation of objects, complex problem solving, or even artistic creation, elementary thoughts are frequently performed without conscious awareness. However, the question of why reasoning is especially related to spatial rather than to visual processes still remains open. To see why this could be the case, it is helpful to compare the spatial account of reasoning with what is claimed about visual mental imagery. On the one hand, nowadays almost everyone believes that the visual system is a distinct input-module of the mind. But it has been frequently argued that visual mental imagery, a purely mental process, involves the resources of the visual module for the purposes of thinking and reasoning. Regardless of the fact that the reported "facts" at least partially contradict this theory, the acceptance of overlapping systems for perception and imagination has important consequences. Clearly, it implies that the visual system cannot be said to be modular in the sense of Fodor (1983). De facto, such a system is not informationally encapsulated, since the characteristics of visual perception have an impact on the imagination process. To insist on the modularity axiom leads to problems for the visual theory of reasoning and vice versa, since there are apparently some mechanisms that are shared by visual mental images and visual perceptions. One possible resolution might be that both processes simply share the same central resources. An alternative explanation is that both perception and imagination have access to and make use of a distinct model — a central spatial system or module. Such a module certainly receives input from visual, auditory, vestibular, and somatosensory perceptual systems, but does not overlap with these subsystems. Thereby, the system could be seen as encapsulated in the sense of Fodor. Moreover, this spatial module could also interact with modules for higher cognitive functions, such as memory, language, and of greater interest to us — reasoning.

## Consequences for the Development of Computational Spatial Assistance Systems

Leaving all speculations aside, what do the psychological "facts" mean for the development of computational spatial assistance systems? Computer scientists working in the area of diagrammatic reasoning (DR) have argued that image-like representations and reasoning processes are useful in solving problems, ranging from the analysis of molecular structure (Glasgow & Papadias, 1992) to the navigation of robots (Stein, 1994). In the meantime, DR is a well-established area of research that is concerned with how visual information can be represented in diagrams, and how these representations can be used for reasoning. The main characteristic of DR systems is a representational scheme that consists of the most important (salient) visual features and properties of an image. Reasoning based on such analog representations is claimed to be more powerful than traditional propositionally based reasoning (Glasgow, Narayanan & Chandrasekaran, 1995). This approach, however, appears to conflict with theories of qualitative spatial reasoning (QSR). Their proponents argue that qualitative representations of spatial relations, together with an appropriate reasoning engine, are a better way to enable computers to make predictions, diagnoses, and plans when quantitative knowledge is unavailable or leads to computationally intractable inferences (Hernández, 1994; Cohn, 1997). Accordingly, the representation formalism of QSR systems typically consists of incomplete knowledge of the topological relationships between entities, along with imprecise information about distances, directions, and relative sizes. However, the present summary of findings shows that the best way to explore this topic is to investigate psychologically how humans represent and reason about spatial information. The most important corollary from the "facts" is that they are consistent with the theory of mental models. Reasoning is based on the construction, inspection, and variation of spatial representations. These spatial representations are models of the spatial relations among entities, that is, they represent which things are where. In inferential tasks, spatial models are likely to exclude visual detail, to represent only the information relevant to inference. This account agrees with the central assumptions of qualitative spatial reasoning (QSR). But is QSR thus akin to human reasoning? On the one hand, cognitive validity can only be claimed in comparison to other approaches that arise with different predictions (Knauff, 1999). In the light of the reported results, QSR indeed appears to be cognitively more plausible than DR accounts of reasoning, since the main assumption that visual representations are essential to inference is not supported by the reported findings. A word of caution, however, is that the visual and spatial nature of representations in reasoning depends on the nature of the problem. Reasoning that we explored might elicit spatial representations, but reasoning with other problems might

elicit visual images in addition. At the workshop a number of other consequences for computational systems will be discussed. In particular, I will be talking about what the findings mean for the externalization of information to support human reasoning.

## Acknowledgments

## References

Cohn, A. G. (1997). Qualitative spatial representation and reasoning techniques In *KI-97: Advances in Artificial Intelligence* (pp. 1-30). Berlin: Springer.

De Soto, L.B., London, M., & Handel, M.S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 2, 513–521

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fangmeier, T., Knauff, M. & Ruff, C.C. (submitted). The Neural Correlates of Logical Thinking: An Event-Related fMRI Study. *Manuscript submitted for publication* .

Glasgow, J., & Papadias, D. (1992). Computational imagery. *Cognitive Science,* 16, 355-394.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: Cambridge University Press.

Hernández, D. (1994*). Qualitative representation of spatial knowledge*. Berlin: Springer.

Kerr, N.H. (1983). The role of vision in "visual imagery" experiments: Evidence from the con-genitally blind. *Journal of Experimental Psychology: General,* 112, 265-277.

Knauff, M. (1999). The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. *Spatial Cognition and Computation, 1,* 261–290.

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience, 4,* 559–573.

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition, 30,* 363–371.

Knauff, M., & May, E. (2004). Visual imagery in deductive reasoning: Results from experiments with sighted, blindfolded, and congenitally totally blind persons. *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Knauff, M., & May, E. (submitted). The role if visual imagery in deductive reasoning: Results from experiments with sighted, blindfolded, and congenitally totally blind persons. Manuscript submitted for publication.

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Imaging deductive reasoning: A functional MRI study. *Cognitive Brain Research, 13,* 203–212.

Knauff, M., Rauh, R., & Schlieder, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 200–205). Mahwah, NJ: Lawrence Erlbaum Associates.

Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.

Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C. & Strube, G. (under revision). From Preferred To Alternative Mental Models In Spatial Reasoning . *Spatial Cognition and Computation.*

Ruff, C.C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: common and distinct neuronal processes. *Neuropsychologia, 141, 1241-1253.*

*Stein, L. A. (1994). Imagination and situated cognition. Journal of Experimental* and Theoretical Artificial Intelligence, 6, 393-407 [reprinted in: K. M. Ford, C. Glymour, & P. J. Hayes (Eds.) (1995), Android Epistemology (pp. 167-182). Cambridge, MA: AAAI Press / MIT Press.