

# AI and Mental Imagery

**Samuel Wintermute**

University of Michigan  
2260 Hayward St.  
Ann Arbor, MI 48109-2121  
swinterm@umich.edu

**Scott D. Lathrop**

United States Military Academy  
EECS  
West Point, NY 10996  
scott.lathrop@us.army.mil

## Introduction

Vision and space are prominent modalities in our experiences as humans. We live in a richly visual world, and are constantly and acutely aware of our position in space and our surroundings. In contrast to this seemingly precise awareness, we are also able to reason abstractly, use language, and construct arbitrary hypothetical scenarios.

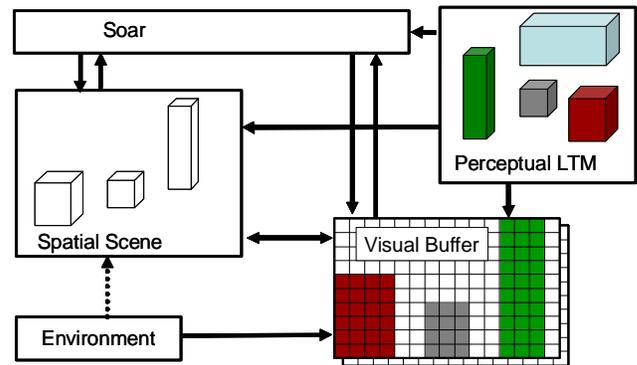
In this position paper, we present an AI system we are building to work towards human capability in visuospatial processing. We use mental imagery processing as our psychological basis and integrate it with symbolic processing. To design this system, we are considering constraints from the natural world (as described by psychology and neuroscience), and those uncovered by AI research. In doing so, we hope to address the gap between abstract reasoning and detailed perception.

## Constraints from AI and Psychology

Historically, one of the most prominent approaches to AI has been symbol processing. While purely symbolic AI has its weaknesses, it has some very important strengths. Symbols allow for very general reasoning and can be composed together to create arbitrary hypothetical situations (Newell, 1990). Humans also exhibit the ability to create arbitrary situations and since symbols are a good AI answer for this capability, we take this as a constraint on our system: it must use symbolic reasoning. Specifically, we are pursuing our research in the context of the Soar cognitive architecture (Laird, 2008), which includes symbolic processing.

Symbolic AI systems typically use qualitative reasoning, where a higher-level representation of the continuous world is reasoned over, rather than precise information as might be provided by the senses. Much work in AI has focused on finding appropriate qualitative representations of space, but this work has led to the poverty conjecture of Forbus et al. (1991), that “there is no purely qualitative, general-purpose representation of spatial properties”. If this is true, it places another constraint on our system: it must employ a non-qualitative representation of space.

Looking to psychology, a relevant area of study is mental imagery (Kosslyn, 2006). We have been



**Figure 1. The SVS system**

investigating two forms of imagery, spatial and visual. In both forms it appears the brain activates its perception system from the top down, imagining objects in the same systems that we normally associate with perception. Humans seem to have specialized systems to handle spatial and visual information, and imagery brings these systems under the umbrella of cognition, since they are used for more than simply translating sensory information into higher-level representations. This is another constraint in our system: it must model human spatial and visual imagery by including representations and specialized forms of processing associated with these forms of imagery.

## The SVS Spatial / Visual System

Soar+SVI (Soar Spatial and Visual Imagery; Lathrop and Laird, 2007; Lathrop, 2008) is a system created to study spatial and visual imagery with Soar, and SRS (Spatial Reasoning for Soar; Wintermute and Laird, 2007, 2008) is a system created to explore problem solving with a spatial representation, focusing on the translation between the spatial and symbolic layers. These systems are being combined and improved in a new system, Spatial and Visual System or SVS (Figure 1), an extension to Soar.

There are two short-term and one long-term memory in SVS. The two STMs are for visual information (roughly corresponding to the visual cortex), and spatial information (roughly corresponding to a region in the parietal cortex). The Visual Buffer is retinotopically mapped, and two-dimensional. It represents strictly visual information, such as color and exact shape. The retinotopic brain structure corresponds to a depictive structure in the computer (a bitmap image), where empty space is represented

explicitly. The Spatial Scene is three-dimensional, and could be inferred from any modality, but we assume that it is inferred from the Visual Buffer during perception. In virtual environments, it is possible to connect the Spatial Scene directly to the environment, as those environments typically use spatial encodings natively. SVS also includes a perceptual LTM, for both spatial and visual information. All of these memories communicate with Soar via a qualitative interface: no low-level spatial or visual information is present there, only high-level information such as object identities and relationships between objects.

The previous systems are subsets of SVS, and we have reported several results in studying them. The constrained, representation-specific processing in the spatial and visual systems provides a functional advantage and is more efficient than processing the same information symbolically (Lathrop, 2008). The use of imagery can allow complicated reasoning processes such as path planning to be split between abstract symbolic reasoning processes and precise spatial processes (Wintermute and Laird, 2007), making an overall system that is both general across problems and precise within problems. Spatial imagery also allows symbolic AI to address problems involving fine-grained continuous motion (Wintermute and Laird, 2008). A common theme in all of this work is that a system with imagery is able to symbolically compose hypothetical scenarios, which can then be precisely interpreted by using imagery (e.g., “What if I tried to move around this obstacle?” “If the enemy was sitting by the hill, could my teammate see him?”).

For example, an agent operating with a teammate in an environment with obstacles and adversaries may use imagery to determine if its teammate is in a good position to over watch an approaching enemy (Lathrop, 2008). To do this, Soar formulates a symbolic description of the hypothesized position of the enemy and teammate, which is then interpreted by the imagery system. Soar can then query the imagery system for qualitative implications of the situation, such as “Does the region viewed by my teammate intersect the enemy?”.

### Future Work

An advantage of examining visuospatial processing from an imagery standpoint is that we can make progress without having to address every problem typically found in vision research. It is much easier to derive a visual representation of a known object in a known position than it is to identify an unknown object and infer its position. As the processes and representations used in mental imagery are shared with perception, studying imagery in AI should aid the study of computer vision. In particular, creating an imagery system requires us to determine what is and is not a sufficient system for representing and using visuospatial knowledge. We hope that this will further constrain the vision problem, aiding research in that area. Similarly, as humans are able to solve the same problems our system is addressing, exploring the details of what architecture is

needed to perform this type of non-logical reasoning can aid the further development of psychological theories.

To further these goals, we are working on extending our system towards robotics. We believe that mental imagery can provide a key link in robotics systems attempting to incorporate a full range of capability, from the sensor/effector level to the cognitive level. Pursuing this presents many scientific and engineering challenges. Most importantly, methods are needed to translate common robotic sensory information into spatial objects or references to prototypical objects in LTM, which can then be retrieved and used in reasoning.

In addition to developing SVS as an AI system, we have long-term plans to extend SVS to model the details of perceptual attention. This capability should allow SVS to serve more directly as a psychological model, since its results could be matched against human data. In addition, this kind of modeling should force the system to encompass a theory of the timing of object recognition, which will move it closer to addressing the mechanisms of object recognition.

### Conclusion

We have been working to integrate a naturally-inspired component, mental imagery, with an existing AI system, Soar. This integration has increased the capabilities of the AI system, and has opened up interesting research directions in both AI and psychological modeling.

### References

- Forbus, K.D., Nielsen, P., Faltings, B. Qualitative Spatial Reasoning: the CLOCK Project. *Artificial Intelligence* 51, 1991.
- Kosslyn, S., Thompson, W., Ganis, G., *The Case for Mental Imagery*. Oxford University Press, 2006.
- Laird, J. E., Extending the Soar Cognitive Architecture, Artificial General Intelligence Conference, 2008.
- Lathrop, S. D., “Extending Cognitive Architectures with Spatial and Visual Imagery Mechanisms,” PhD Thesis, University of Michigan, 2008.
- Lathrop, S. D., and Laird, J.E., Towards Incorporating Visual Imagery into a Cognitive Architecture, International Conference on Cognitive Modeling, 2007.
- Newell, A. *Unified Theories of Cognition*. Harvard University Press, 1990.
- Wintermute, S., and Laird, J.E., Predicate Projection in a Bimodal Spatial Reasoning System, AAAI Conference on Artificial Intelligence, 2007.
- Wintermute, S., and Laird, J.E., Bimodal Spatial Reasoning with Continuous Motion, AAAI Conference on Artificial Intelligence, 2008.