In Proceedings of ICCM - 2007- Eighth International Conference on Cognitive Modeling. 267 - 272. Oxford, UK: Taylor & Francis/Psychology Press.

Modeling Memories of Large-scale Space Using a Bimodal Cognitive Architecture

Unmesh Kurup (kurup@cse.ohio-state.edu)
B. Chandrasekaran (chandra@cse.ohio-state.edu)
Computer Science and Engineering, 2015 Neil Ave
The Ohio State University, Columbus, OH 43210

Abstract

We present an extension to biSoar, a bimodal version of the cognitive architecture Soar, by adding a bimodal version of *chunking,* Soar's basic learning mechanism. We show how this new biSoar is a useful tool in modeling cognitive phenomena involving spatial or diagrammatic elements by applying it to the modeling of problem solving involving large-scale space, such as way-finding. We suggest how such models can help in identifying variables to control for in human subject experiments.

Introduction

Cognitive architectures are of central interest in cognitive modeling since such architectures are directly useful in building cognitive models. The advantages of a general purpose architecture such as Soar or ACT-R to model and explain a variety of cognitive phenomena are well-known. However, these architectures are all based on a view of the cognitive state being symbolic or more precisely, predicate-symbolic. In this view, the agent's knowledge, goals etc are represented in terms of symbol structures that describe the world of interest in terms of properties of and relations between individuals in the world. We have argued that this view of cognitive state is too restrictive and fails to take adequate account of the role played by perceptual representations in thinking (Chandrasekaran, 2006). We have proposed that cognitive state should be viewed as multi-modal where, in addition to the traditional symbolic component, the cognitive state has several perceptual components and a kinesthetic one. The multi-modal view proposes a more involved role for perception where the perceptual systems, in addition to their role as transducers, also provides representations and processes to the cognitive process. Such a multi-modal state can support an agent experiencing the world multi-modally such as when having mental images in one or more modalities. One of the tasks of a research program that is based on this multi-modal view is to explore the consequences of multi-modal cognitive state for all components and mechanisms of a cognitive architecture.

In particular, we need to examine the implications of multi-modality for components such as working memory, LTM and I/O and for control and learning mechanisms. As a first step towards constructing such a multi-modal architecture, we built biSoar (Kurup & Chandrasekaran, 2006), which is a bimodal augmentation of the Soar architecture, where the two modes are the traditional symbolic component and a visual (diagrammatic) component. This limitation to bimodality has several advantages. First, intuitions about various issues related to multi-modality may be honed by investigating this limited version. Second, in problem solving, the most common and useful perceptual mode is the limited visual version involving diagrams. Soar was chosen for reasons of convenience but we think that many of the ideas in biSoar can be extended to other symbolic architectures such as ACT-R. However, Soar also has unique mechanisms such as chunking as a core learning mechanism, an issue that will be a focus of the current paper. As an aside, the visual component does not represent all aspects of mental imagery such as the visualization of faces but is restricted to diagrams. In addition to diagrams being common in problem solving, the focus on diagrams has the advantage of simplicity while retaining several of the challenges of bimodality that we wish to address.

Currently, both working and long term memories are bimodal in biSoar. biSoar agents are able to create, modify and delete diagrammatic objects from WM as well as extract various relations that exist between objects in this memory. However, among the issues not addressed is how diagrammatic information gets into long term memory. Phenomenologically, it seems clear that memory is capable of recalling perceptual knowledge and experience to a more or less degree of fidelity. It seems plausible that in the course of learning, learning mechanisms transfer to long term memory not only symbolic information from working memory but diagrammatic information as well. In traditional Soar, there is only one learning mechanism, chunking. So it seemed natural to us to investigate how chunking can be expanded to learn bimodally. An empirically observed feature of many spatial memories is that spatial details are often simplified (Tversky, 2000). So, a challenge for bimodal chunking is the degree to which such simplification is an intrinsic architectural feature.

A satisfactory account of bimodal learning could make an architecture with such a capability an effective medium for modeling cognitive phenomena involving spatial or diagrammatic elements. We explore the possibilities of biSoar for such modeling by applying it to the task of modeling phenomena involving the representation of and reasoning about large-scale space. We build biSoar models of problem solving in two spatial reasoning tasks that have been well studied: simplification in recalled routes and distortions in geographic recall. Such modeling can be a valuable tool for exploring the space of explanations for spatial phenomena. For each task, we create multiple models and describe how each one suggests a different explanation for the observed phenomena. We indicate how a candidate explanation can in turn suggest variables to control for in human subject experiments.

Multi-modal Cognitive Architectures

The traditional approach to cognition and problem solving can be best described "predicate-symbolic"; that is, the knowledge and goals of an agent are represented as a set of entities, and relations (predicates) that hold between these entities. Problem solving proceeds by the application of rules of inference to these predicates. The role of the perceptual system is to give the agent information about the external world, and the role of the action system is to make changes to the world. The output of the perceptual systems, in this view, is in the form of predicate-symbolic representations. Our alternative proposal calls for a much greater role for an agent's perceptual system in cognition. Here, the agent has representations and processes that are characteristic to the individual modalities and cognition is an activity that involves all of them. The perceptual system as whole still give information about the external world, but aspects of the system are part of central cognition, independent of input from the external world.

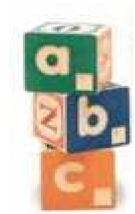
To create biSoar (Kurup & Chandrasekaran, 2006), a general-purpose cognitive architecture, Soar (Laird *et al.*, 1987) was augmented with the Diagrammatic Reasoning System (DRS) (Chandrasekaran *et al.*, 2004), a domain independent system for representing diagrams. In DRS, diagrams are represented as a configuration of points, curves and regions. That points may refer to the location of cities or that regions represent states in a map, is task-specific knowledge that is part of Soar. This allows DRS to be used in multiple task domains without any modifications. DRS also provides a set of perceptual and action routines that allows Soar to create, and modify a diagram and to extract relations between diagrammatic objects from the diagram. By the addition of the capabilities of DRS, Soar's cognitive state and long-term memory that were exclusively predicate-symbolic, now become bimodal.

Cognitive State in Soar

Soar's representations are predicate-symbolic. The cognitive state in Soar is represented by the contents of Soar's WM and operator, if any, that has been selected. Fig 1(b) shows Soar's cognitive state representation of the blocks world example in 1(a).

Cognitive State in biSoar

The cognitive state in biSoar is bimodal – it has both symbolic and diagrammatic parts. Fig 2 shows the bimodal representation of the world depicted in Fig 1(a). Working memory is biSoar is represented as a quadruplet, with each Identifier, Attribute, Value triplet augmented with a diagrammatic component in DRS that represents the visualization (metrical aspect) of the triplet. Since not all triplets need to be (or can be) visualized, the diagrammatic components are present only as needed. States represent the



(a)

Working Memory:
Block (A), Block (B), Block (C), On (A,B), On (B,C)

Selected Operator: None

(b)

Fig 1: (a) Blocks World and (b) Soar's representation of the world in (a).

current or potential future state of interest in the world and the symbolic and the diagrammatic part may represent related or distinct aspects of the world. However, the diagrammatic representation is "complete" in a way that the symbolic representation is not. For example, from the symbolic representation alone it is not possible to say without further inference whether A is above C. But the same information is available for pick up in the diagram with no extra inference required. This has advantages (for

Working Memory:
Block (A), Block (B), Block (C), On (A,B), On (B,C)

Selected Operator: None

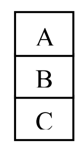


Fig 2: biSoar representation of the world shown in 1(a)

instance in dealing with certain aspects of the Frame Problem) and disadvantages (over-specificity).

From External Representation to Working Memory

When an agent makes use of an external diagram, such as a map, for a specific problem solving task, what he attends to or observes is only relevant parts or aspects of the diagrammatic elements. This selective attention results in simplified versions of the corresponding diagrammatic elements to be present in WM. The mechanism that transforms an external diagrammatic element into a simplified version in WM is part of human perceptual machinery and is needed as an adjunct to biSoar as well. In this paper, we refer to this attention-controlled mechanism as the simplification mechanism. This mechanism is implemented as an A*ttend* method that is part of any routine that interacts with an object in the external world. The A*ttend* method produces the equivalent of the product of attention on aspects of the diagrammatic object. One way to think of A*ttend* is that it is as if parts of the diagrammatic object on which attention is not focused is at a very low resolution resulting in the loss of many of the finer details while still preserving the general spatiality of the object. Fig

3(b) is the output of the A*ttend* operator on the curve in 3(a) where the attention has been focused on just the beginning and end points. Fig 3(d) is the result of A*ttend* on Fig 3(c) where the attention has been focused on the region's broad shape. The result of A*ttend* does depend upon the requirements of the task because that determines the aspects to which attention was paid to in the diagram. But simplification in this manner is architectural because it happens irrespective of the task or the domain.

Bimodal LTM

There are two questions that have to be answered in an implementation of Long Term Memory (LTM) – how are elements put into LTM (i.e., learned) and how are elements retrieved from LTM. In the case of Soar the answers to these two questions are chunking for learning and a matching process that matches the LHS of a LTM rule to WM for retrieval.

Chunking - Chunking simply transfers the relevant contents of WM to LTM. In the case of biSoar, chunking transfers to LTM the simplified versions of the relevant external diagrammatic elements present in WM.

Matching - In the case of Soar the retrieval process is straightforward because matching (or even partial matching when variables are present) symbols and symbol structures to each other is an exact process; either they match or they don't. When the cognitive state is bimodal, WM has metrical elements in addition to symbols. Matching metrical elements to each other (say a curve to another curve) is not an exact process since two metrical elements are unlikely to be exactly the same. Matching metrical elements would require a different approach like a non-exact process that can match roughly similar elements in a domain-independent manner (since the matching should be architectural). It may also turn out that only calls to perceptual routines are present in LTM while matching metrical elements is a more low-level cognitive process present only in stimulus-response behavior. For now we take the latter approach where the LHS of biSoar rules contain only perceptual calls to the DRS that return symbol structures in addition to symbol structures. We think that this approach can account for many of the diagrammatic learning capabilities that are required in models of cognition except in cases where goal specifications contain irreducible spatial components, such as might be the case in the problem solving of a sculptor. The RHS of a biSoar rule can modify either symbolic or diagrammatic parts of WM.

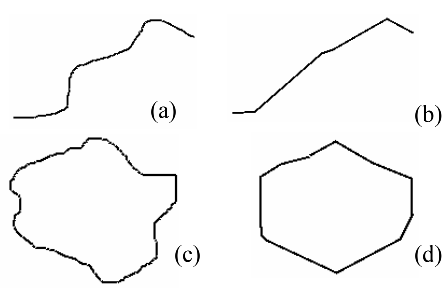


Fig 3: (b) and (d) show the result of applying the visualize operator to (a) and (c) respectively

Representation of Large-Scale Space

In 1948, Tolman (1948) proposed that animals have an internal representation of large-scale space which he called the cognitive map. In 1960, Lynch (1960) produced his seminal study of the environment in which he identified *Landmarks, routes, nodes*, *districts* and *edges* as the features that are important in building a cognitive map. Since then there have been a number of models, both computational and cognitive, that have been proposed to account for a number of phenomena associated with the representation of space. A variety of behavioral/psychological studies have also aided the development of these models by providing a set of characteristics or behaviors that a model should posses.

We believe the use of a general-purpose cognitive architecture such as biSoar can be beneficial in the area of modeling spatial phenomena for three reasons – First, it restricts spatial information to be learned, represented and used within the constraints of a general cognitive architecture. Second, it allows the modeler to be flexible in the strategies and knowledge that they use to model phenomena. Third, it is easier to identify the nature of the explanation (architectural vs. content) because these are explicitly distinguished in such a framework. We use biSoar to model two commonly observed phenomena in spatial reasoning - simplification in recalled routes and distortions in the recall of relations between geographic entities.

Sources of Map Knowledge

Knowledge of large-scale space can come from multiple sources. The most common, of course, being personal experience of navigation in space. We automatically build representations of our environment as we traverse them. A second, and important, source is maps. Our knowledge of large environments, such as the spatial extent and geographical locations of the fifty states, originated from our use of maps. Representations, originating from either source, are combined and modified in various ways during problem solving for various purposes. In this paper, we focus on phenomena involving maps.

The Space of Explanations

When models are implemented in a cognitive architecture as possible explanations for a phenomenon, the behavior of interest can arise from one, or a combination of, two influences:– Architectural and Content where Content can be further sub-divided into Strategy and Knowledge.

An architectural explanation appeals to the specifics of the architecture of the agent to explain the phenomenon of interest. The phenomenon is produced as the result of a process that is automatic and arises out of the architecture, not a deliberative decision by the agent. A phenomenon can also emerge as a result of a particular strategy employed by the agent to solve the given task. This is different from an architectural explanation because the phenomenon is unique to the current task. An agent's behavior can also be seen as arising from its knowledge (or lack thereof) of the task domain and the world. During problem solving, an agent may learn to solve the problem one way due to the knowledge it has at the time. Given more knowledge, the

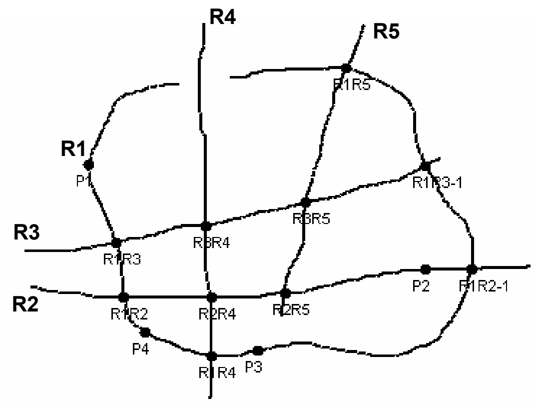


Fig 4: The map for models Simp1 and Simp2. Routes are found from P1 to P2, P4 to R1R3-1 and R1R4 to R1R5

agent might have learned to solve the problem in a different way resulting in different observable phenomena.

In general, a phenomenon can have more than one explanation and it is difficult for an outside observer to decide if the reason for the phenomenon is architectural, strategic or knowledge related without further experimentation. Also, due to the number of free variables and tunable parameters in cognitive architectures, and the fact that they are essentially Turing machines, the ability (or inability) to build a model in the architecture cannot be taken as the final word on whether the explanation offered by the model is correct (or incorrect). Under certain circumstances, however, the inability to build a model in this framework can be taken as a sign that the approach (or strategy) is flawed. More importantly, building cognitive models help us identify the possible sources of a phenomenon. This can in turn be used to develop a series of controlled experiments to decide between the sources.

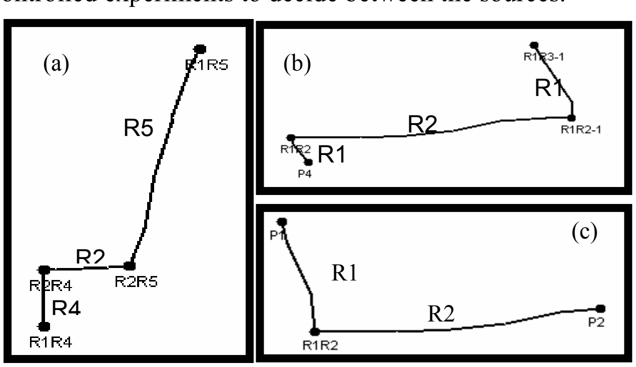


Fig 5: Routes found by Simp1 from the map in Fig 4. (a) R1R4 to R1R5. (b) P4 to R1R3-1. (c) P1 to P2

Task 1 – Simplification in Route Recall

Curves recalled from spatial memories, whether they are rivers in Paris or routes by cab drivers rarely preserve their exact curvature or their orientation to each other and to other landmarks (Tversky, 2000). Details in a curve such as the actual angles at intersections are lost and route curvature is usually straightened. In this paper, we refer to this phenomenon as *simplification*. We explore how this phenomenon can arise from the architectural features of biSoar. In particular, we explore whether the chunking of the simplified diagram in WM (represent only that to which attention was paid) is enough to explain the emergence of simplification in recalled maps.

Model 1

The agent (referred to as Simp1) is given the task of finding various routes in the map shown in Fig 4. Fig 5 shows the result of route-finding for certain locations from the map. The route-finding strategy used is a simple one in which the agent finds the routes on which the current point lies, finds the next point along all possible directions, calculates the Euclidean distance to the destination from each point and picks the one with the lowest value. The critical step in the strategy is the step where, once the next point has been selected, the agent notes the route from the current point to the selected next point paying attention to only the starting and ending points of the route. This results in a representation of the route that is simplified according to the attentional demands of the task, in WM. When Soar's chunking mechanism learns from the resolution of the sugboal, it learns this simplified representation from WM.

Model 2

A new agent (Simp2) is created and given the same task as Simp1. Simp2's strategy is the same as Simp1's except that Simp2 chooses to pay attention to only the locations of important intersections and the names of the routes they lie on. During recall, Simp2 recalls these locations and connects them using straight lines. Fig 6 shows routes recalled between the same locations as in Fig 5.

Discussion

The two models (represented by the two agents Simp1 and Simp2) indicate two different explanations for the simplification phenomenon. The simplified routes recalled by Simp1 are the result of an architectural feature of biSoar – bimodal chunking. Depending on which aspects of the routes that attention was paid to, Simp1 chunks a simplified version of the original route. Simp2 on the other hand, does not even bother trying to chunk the spatiality of the routes. Instead, it learns the locations of important intersections and the routes they are on and connects them with straight lines during recall. As mentioned before, the ability to create these models does not automatically suggest that either (or both) explanation is the definitive source of the simplification phenomenon. There could be other as yet unwritten models that might turn out to be, in fact, right. However, these models do suggest that one variable to control for is whether subjects are recalling only locations or both locations and routes. One way to do this would be to have a particularly attention grabbing feature on one of the curves (maybe a loop or sudden change in direction).

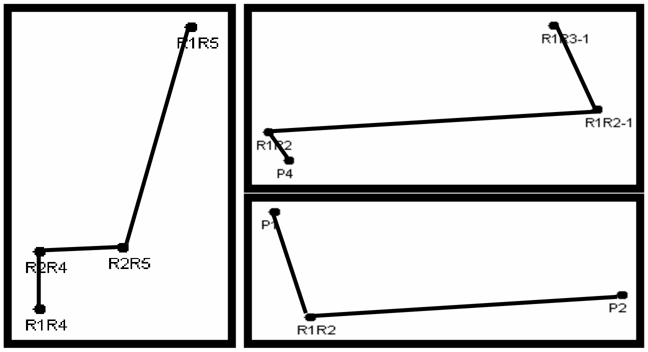


Fig 6: Routes found by Simp2 from the map in Fig 4. (a) R1R4 to R1R5. (b) P4 to R1R3-1. (c) P1 to P2

Task 2 – Distortion in Geographic Recall

According to Stevens and Coupe (1978), when subjects were asked about the relation between San-Diego and Reno, most answered that San-Diego was to the west of Reno even though in reality, Reno is west of San Diego. They go on to suggest that this result indicated two things – one, that the cognitive map was unlikely to be a faithful metrical representation and two, that the representation was hierarchical in nature, the hypothesis being that since the subjects did not have any information about the relationship between SD and Reno they went up the hierarchy and compared the containing regions – California and Nevada. Since California is to the West of Nevada, it followed that SD was to the west of Reno.

We built three different models of problem solving for this task. Model 1 is of an agent that has a single simplified metrical representation of California and Nevada in LTM (and WM) like in Fig 7 (a). In this particular example San Diego to the West of Reno, but an agent that paid particular attention to these cities may have a metrical representation with the cities in their correct relationship to each other. Model 2 has symbolic information in LTM that San Diego is South of San Francisco and that Reno is East of San Francisco. It constructs a diagram (Figure 7(b)) in WM using this information and extracts the (wrong) answer from the diagram. Model 3 has symbolic information in LTM that San Diego is in California, Reno in Nevada and that California is to the West of Nevada. This information is used to construct a diagram (Figure 7(c)) and the (wrong) answer extracted from it.

Discussion

The variety of models in Task 2 exhibit biSoar's flexibility in modeling spatial phenomena. Each model provides a different explanation and, in essence, suggests a separate control variable. For example, in Model 2, the explanation is that subjects use a specific strategy – that of comparing the location of the target cities to a common city and inferring the relationship from that knowledge. This strategy can be controlled for by using artificial maps (as Stevens and Coupe do in their original paper) that do not provide this extra information. Thus, models in biSoar have a straightforward mapping to issues to control for and building these models provides a natural way of discovering

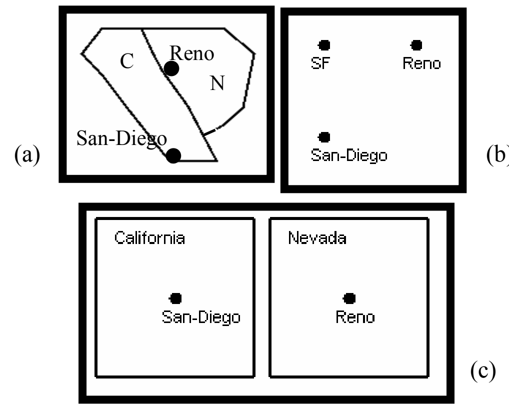


Fig 7:(a) Map of SW U.S. in LTM (& WM) of Model 1. (b) & (c) are diagrams in WM constructed by Models 2 & 3

these issues. Of course, the experimenter is free to simply think of various explanations without modeling in biSoar, but the advantage is that it provides additional constraints and restricts the experimenter to those explanations that are cognitively possible. The disadvantage is that we do not know of any systematic way of generating these models/variables. Certain heuristics such as "look for at least one explanation from each possibility in the explanation space" can suggest lines along which the model builder/experimenter may approach the problem.

Related Work

Soar – Lathrop and Laird (2006) report on progress in their work on expanding Soar to include a perceptual representation and reasoning system. There is at least one important theoretical distinction between their work and ours. Our work is based on the assumption that all aspects of the agent's architecture including the cognitive state, memory, learning etc, are multi-modal and that during problem solving Soar can seamlessly access representations across all modalities. Lathrop and Laird take a different approach, one in which the perceptual system is part of the total cognitive system, but outside of high-level cognition. This means that perceptual representations can be accessed only through the input/output system and access to them is restricted to the input and output phases of Soar's decision cycle. In practice, the implementations are very similar and we believe their system can model most of what we do, including the visualization of information and subsequent extraction of the desired spatial relationship as in Model 3. However, they do not as yet have a theory of automatic learning (what we refer to as bimodal chunking) for the visual part, which provides the basis for an architectural explanation of phenomena such as simplification.

ACT-R – ACT-R or Adaptive Control of Thought – Rational (Anderson *et al.*, 2004) is a general cognitive architecture whose goal is to model all aspects of high-level human cognitive activity. However, there are no reports on any work in augmenting ACT-R's cognitive state to be multi-modal. Certain related work such as ACT-R/S (S for spatial) (Harrison & Schunn, 2002) augment ACT-R with representations for immediate space and object shapes for manipulation but there is no claim to a diagrammatic component that unifies experience whether from memory or perception.

Other Work – There have been a number of non-cognitive architecture oriented proposals for spatial representation and learning, notably the Spatial Semantic Hierarchy or SSH (Kuipers, 2000). The SSH is a multi-layered theory, that represents its knowledge of space at multiple levels – control, causal, topological and metrical, with the information at one level building on what was learned at the next lower level (except in the case of the metrical level.) In its current avatar, biSoar encompasses the topological and metrical levels of SSH. The representational and problem solving capabilities of biSoar and SSH with regards to topological information are similar. The real difference is at the metrical level. SSH proposes a few ways in which 2-D metric information may be represented but biSoar, and in particular, DRS provide a concrete representational format for metric information. Further, biSoar creates, modifies and inspects this information during problem solving making DRS an integral part of the problem solving process.

Other models include Absolute Space Representation (ASR) (Jefferies & Yeap, 2001) and MIRAGE (Barkowsky, 2001). Both combine models of representation with a metrical representation that has aspects of DRS.

Since SSH, ASR and MIRAGE are all intended to model spatial representation and reasoning, they lack the flexibility of a general cognitive architecture that biSoar provides.

Concluding Remarks

We have presented a proposal for bimodal learning within the existing learning mechanism in Soar, chunking, and shown how building models of spatial representation and reasoning within this architecture can help in the design of experiments. We believe that a bimodal architecture augmented with bimodal chunking can be an useful vehicle in exploring the nature of the human cognitive map.

Two additional details need to be satisfactorily addressed for a measure of closure in this direction of research. The first relates to biSoar's rule matching process mentioned earlier, where elements in LTM rules are matched against structures in WM. It is not yet clear how to match diagrams on the LHS of rules to the diagrammatic part of WM. Second, the processes involved in composing diagrammatic elements from different LTM rules in WM according to the needs of the current goal. For example, subjects may remember the border of Texas using multiple diagrams - one consisting of a simplified overall view, another representing the "top hat" part and a third representing the coastline. During recall, the diagram is constructed by integrating these overlapping or locally inconsistent images with the aid of task-specific knowledge.

We believe that the idea of simplification that is presented extends naturally to other memories such as semantic and episodic memories. Further, even though it is presented in the context of Soar, the general ideas are likely to be applicable to other symbolic architectures like ACT-R.

Acknowledgements

Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Barkowsky, T. (2001). Mental processing of geographic knowledge. In D. R. Montello (Ed.), *Spatial information theory - foundations of geographic information science* (pp. 371-386). Berlin: Springer.

Chandrasekaran, B. (2006). *Multimodal cognitive architecture: Making perception more central to intelligent behavior.* Paper presented at the 21st National Conference on Artificial Intelligence, California.

Chandrasekaran, B., Kurup, U., Banerjee, B., Josephson, J. R., & Winkler, R. (2004). *An architecture for problem solving with diagrams.* Paper presented at the *Diagrammatic Representation and Inference conference*.

Harrison, A. M., & Schunn, C. D. (2002). *Act-r/s: A computational and neurologically inspired model of spatial reasoning.* Paper presented at the 24th Annual Meeting of the Cognitive Science Society, Fairfax, VA.

Jefferies, M. E., & Yeap, W. K. (2001). *The unity of global representations in a cognitive map.* Paper presented at the International Conference on Spatial Information Theory: Foundations of Geographic Information Science.

Kuipers, B. J. (2000). The spatial semantic hierarchy. *Artificial Intelligence, 119*, 191-233.

Kurup, U., & Chandrasekaran, B. (2006). *Multi-modal cognitive architectures: A partial solution to the frame problem.* Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1-64.

Lathrop, S., & Laird, J. E. (2006). *Incorporating visual imagery into a cognitive architecture: An initial theory, design and implementation* (No. CCA-TR-2006-01).

Lynch, K. (1960). *The image of the city*. Cambridge: MIT Press.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology, 10*, 422-437.

Tolman, E. C. (1948). Cognitive maps in rats and man. *Psychological Review, 55*, 189-208.

Tversky, B. (2000). Levels and structure of spatial knowledge. In R. Kitchin & S. Freundschuh (Eds.), *Cognitive mapping: Past, present, and future* (1 ed., pp. 24-43): Routledge.