A Bimodal Cognitive Architecture: Explorations in Architectural Explanation of Spatial Reasoning

B. Chandrasekaran and Unmesh Kurup

Department of Computer Science and Engineering
The Ohio State University
Columbus, OH 43210
{chandra|kurup}@cse.ohio-state.edu

Abstract

Research in Psychology often involves the building of computational models to test out various theories. The usual approach is to build models using the most convenient tool available. Newell has instead proposed building models within the framework of general-purpose cognitive architectures. One advantage of this approach is that in some cases it is possible to provide more perspicuous explanations of experimental results in different but related tasks, as emerging from an underlying architecture. In this paper, we propose the use of a bimodal cognitive architecture called biSoar in modeling phenomena in spatial representation and reasoning. We show biSoar can provide an architectural explanation for the phenomena of simplification that arises in experiments associated with spatial recall. We build a biSoar model for one such spatial recall task – wayfinding, and discuss the role of the architecture in the emergence of simplification.

Introduction

A common research approach in psychology is to start with a body of observations about human behavior or performance in some general task and propose a model that explains the performance. In recent years, such models are often computational and may be based implicitly on a general view of cognition, but the models themselves are task-specific. Alternatively, Newell proposed that the field move towards unifying explanations by using general cognitive architectures as the framework for model-building (Newell 1990). The same architecture would be used to build models for different phenomena and while the models would still be task-specific, the underlying architecture would be general. Newell argued, by using the example of "Salthouse Twentynine," (Newell 1990) how such unification provides more perspicuous explanations of experimental results in different but related tasks, as emerging from an underlying architecture that adapts itself to the variety of tasks. Newell and his associates developed Soar as such a general architecture and Anderson and his associates developed ACT-R for similar purposes (Anderson, Bothell et al. 2004). These architectures as a rule focus on what is often called "central cognition." Perception and motor systems that necessarily would be involved in physical activities such as navigating in real physical space would be external to central cognition.

Over the recent decades a substantial body of literature has grown around representation and use of knowledge of large-scale space, motivated both by a desire to understand and account for human performance on navigation, memory for routes and related tasks, and by the need to provide robots with suitable tools for navigation. Some of this research uncovers certain observable properties of human performance (such as certain kinds of systematic errors) with implications, some times spelt out and other times not, for forms of underlying spatial knowledge (Tversky 1993). Some are specific proposals for knowledge representation (Kuipers 2000) while others provide specific computational models for some of the phenomena (Barkowsky 2001).

We propose the use of a general-purpose architecture to model the representation of and reasoning about large-scale space. However, the architecture we propose to use for such modeling is based on our earlier proposal for a fundamental change in the nature of the cognitive architecture: that it be augmented with perceptual components that play a significant role in central cognition. We describe an architecture called biSoar, which embodies our proposed augmentation of a visual/spatial component to Soar. We show how biSoar can provide an architectural explanation for the simplification phenomenon that is common in spatial reasoning. Simplification is a phenomenon where the spatial details (not relevant to the task) of entities are left out during spatial recall. Simplification can be commonly seen in sketch maps that people draw to provide directions from one location to another. In such maps, routes are often straightened, regions that are nearby are combined into a single region (or left-out altogether), routes that intersect at angles are drawn perpendicular to each other, all while achieving the goal of conveying the information required to reach the destination. We believe simplification plays a recurring role in spatial reasoning and is an underlying factor in explaining multiple phenomena because it restricts the information that an agent has available to it whether it is in navigating or map building or spatial recall.

In general, when models are implemented in some cognitive architecture as possible explanations for a phenomenon, the behavior of interest can arise from one, or a combination, of three influences:– Architectural, Strategy and Knowledge. Architecture is task-independent, Strategy is task-dependent, and Knowledge is even more so in that agents using the same strategy may differ in specific pieces of knowledge.

- Architectural Explanations – An architectural assumption appeals to the specifics of the architecture of the agent to explain the phenomena of interest. For instance, one way an agent can learn is through an automatic mechanism that puts away the knowledge that the agent attended to while solving a sub-problem. This compiled knowledge can be used the next time the agent encounters the same sub-problem. This process is automatic and a feature of the architecture, not a deliberative decision by the agent.
- Strategy Explanations – A phenomenon can also emerge as a result of a particular strategy employed by the agent to solve the given task. Using the learning example from before, an agent can also learn due to a deliberate decision to put away knowledge while solving a particular task. This is different from an architectural explanation because learning in this scenario is specific to the current task while, in an architectural explanation, learning is automatic and happens regardless of the task or problem being solved.
- Knowledge Explanations – An agent's behavior may also arise from the presence or absence of specific items of domain knowledge. If the agent doesn't know there's a bridge at a certain location, it is not going to generate routes that could make use of the bridge.

In general, phenomena can have more than one explanation. In a wayfinding example for instance, simplification can arise due to the strategy used. Consider an agent finding a route from Columbus to Buffalo. It looks at the map in front of it and remembers the relative locations of important cities along the route. The route found by the agent would be from Columbus to Cleveland, to Erie and then to Buffalo. When asked to recall the route, the agent retrieves from memory the relative locations of the important cities and draws straight lines connecting them. To an outside agent simplification has occurred but, without further experimentation, it cannot decide if the reason for simplification is architectural or strategic. Due to the number of free variables and tunable parameters in cognitive architectures, 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 the general cognitive architecture framework can be taken as a sign that the approach (or strategy) is flawed. More importantly, building models gives us a list of possible explanations for the phenomenon. This list can then be used to develop a series of controlled experiments that can decide between the various explanations.

In the next section we describe some of the existing literature on the representation of large-scale space. Following that we briefly describe biSoar. We look at our initial experimental results in writing a biSoar model that exhibits the phenomenon of simplification during wayfinding. We compare our proposal with an existing task-specific proposal, namely SSH, and conclude by discussing some future directions for this research.

Representation of Large-Scale Space

In 1948, Tolman proposed that animals have an internal representation of large-scale space which he called the cognitive map (Tolman 1948). Though he referred to it as a map only in the functional sense, early speculation involved around whether this representation was really a map (a global metrical representation) or merely a collection of representations (only a minority of which were metrical). An overview of the debate is presented in (McNamara 1986). In 1960, Lynch produced his seminal study of the environment and its features that are important in building a cognitive map (Lynch 1960). Lynch identified *Landmarks* – salient cues in the environment such as distinctive buildings, *routes* such as roads, rails and even bike pathways that connect various landmarks, junctions or intersections of routes called *nodes*, *districts* which are implicit or explicit regions of the city, and *edges* that prevented travel, demarcated the different regions and bounded the city itself.

The longest standing model of large-scale space representation is the Landmark, Survey, Route (or LRS) model (Siegel and White 1975). LRS theory states that an agent first identifies landmarks in an environment, adds route knowledge between landmarks as he/she traverses the environment and finally adds survey (or configurational) knowledge as the agent becomes familiar with the environment. Once survey knowledge has been added, the agent has the capability to propose novel, previously un-traversed paths between landmarks. In 1978, Stevens and Coupe proposed a hierarchical model of spatial memory to account for distortions in judgments of relative geographical locations (Stevens and Coupe 1978). For example, when subjects were asked if San Diego, CA was to the west of Reno, NV, a number of them said yes even though it was incorrect. Stevens and Coupe hypothesized that subjects stored spatial information about cities and states as hierarchies, and errors in judgment occurred because relation information is not stored at every level and subjects tried to conform the relation between cities to the relation of the corresponding super ordinate political units. Later theories have modified these models in various ways. For example, in 1998 Gilner and Mallot proposed the view-graph theory in which views and egocentric vectors replaced places (view-independent) and allocentric vectors in the cognitive map (Gillner and Mallot 1998). 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.

Knowledge of large-scale space can come from multiple sources. The most common source is, of course, from 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.

biSoar Architecture

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 in terms of properties of and relations between (predicates) individuals in the domain of discourse. 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 as expressed in the action predicates generated by the problem solver. The output of the perceptual systems, in this view, is in the form of predicate-symbolic representations. Beyond providing information in this form, perceptual systems do not participate in the problem solving process, i.e., they are not part of the cognitive architecture and are external modules. 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 a whole still gives information about the external world, but aspects of the system are part of central cognition, independent of input from the external world.

Soar

Soar is an architecture for constructing general cognitive systems (Laird, Newell et al. 1987). Towards achieving this goal, Soar provides representations for short and long-term memory, mechanisms for interacting with the external world, a sub-goaling strategy that is independent of the task and domain and a learning mechanism that allows Soar to learn as a result of success in solving sub-goals. The Soar architecture also provides a rule-based programming language that can be used to program the intelligent agent. Soar's Working Memory (WM) is represented as Identifier, Attribute, Value triplets (Ex: (S1 Object O1) (O1 Color Red)). Long term memory (LTM) in Soar is a collection of rules. Each rule has a condition (*if*) part that is matched to WM. If a match exists WM is changed according to actions specified in the action (*then*) part. There are two kinds of rules – operator proposal and operator application. Proposal rules propose operators. Each operator can be thought of as the next possible step to take in the problem solving process. Application rules apply the actions of the respective operators. Fig 1 shows an example of operator proposal and operator application rules. During problem solving, Soar goes through a series of 5 phases – input, proposal, decision, apply and output.

Operator Proposal: *If* (Color, Red) *then* propose operator to stop car.
Operator Application: *If* operator proposed to stop car, *then* stop car.

Fig 1: Examples of operator proposal and application rules in Soar

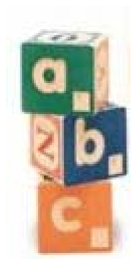
In the proposal phase, all operators that are relevant to the situation (that match against conditions in WM) are proposed. In the decision phase, an operator is selected and the corresponding application rule is executed in the apply phase.

Soar's learning mechanism is called *chunking*. If Soar becomes stuck (called an *impasse*), it creates a sub-goal to try and resolve the problem. For example, if, during the decision cycle, Soar does not know which operator to select, it creates a sub-goal to try and choose an operator. The sub-goal goes away when the impasse that created it is resolved and the information that caused to be resolved is used to create a rule called a *chunk*. The next time Soar is faced with the same problem the chunk is executed instead of re-solving the problem.

To create biSoar, Soar is augmented with a Diagrammatic Reasoning System (DRS). DRS is a domain-independent system for representing diagrams. In DRS, diagrams are represented as a collection of *points*, *curves* and *regions*. The fact that points refer to the location of cities or that regions represent states in a map, is task-specific knowledge that is part of Soar but not of DRS. This allows DRS to be used in multiple task domains without 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, long-term memory etc, that were exclusively predicate-symbolic, now acquire a visual component and hence are bimodal. Our current focus is on diagram-like aspects of the visual component. This allows us to concentrate on the visual reasoning process without distractions from problems that are more involved with image processing. (Chandrasekaran 2004) contains further details about DRS and the perceptual and action routines. (Kurup and Chandrasekaran 2006) contain further details about how DRS is hooked up to Soar.

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 2(b) shows the Soar's cognitive state representation of the blocks world example in 2(a). The world represented by Soar is shown in 2(a).



(a)

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

Selected Operator: None

(b)

Fig 2: (a) Blocks World and (b) Soar's representation of the world in (a).

Cognitive State in biSoar

The cognitive state in biSoar is bimodal – it has both symbolic and diagrammatic parts. Fig 3 shows the bimodal representation of the world depicted in Fig 2(a). Working memory is biSoar is represented as a quadruplet, with each Identifier, Attribute, Value triplet augmented with a diagrammatic component. The diagrammatic component is represented using DRS. It represents the visualization of the triplet. Since not all triplets need to be (or can be) visualized, these components are present only as needed. States represent the current or potential future state of interest in the world and the symbolic and the

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

Selected Operator: None

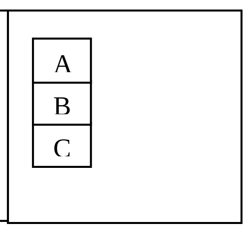


Fig 3: biSoar representation of the world shown in 2(a)

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 instance in dealing with certain aspects of the Frame Problem) and disadvantages (over-specificity). The implications of the use of DR is however outside the scope of the current paper.

To utilize the bimodal cognitive state, Soar's rules in LTM are also suitably modified to be bimodal. Conditions and actions in rules can refer to either symbolic or

Operator Proposal: *If* goal is On(A,B) and In_The_Diagram(A not on B) *then* propose operator to move A on to B.
Operator Application: *If* operator proposed to move A on to B and In_The_Diagram(A and B are clear) *then* In_The_Diagram(move A on to B).

Fig 4: Examples of operator proposal and application rules in biSoar

diagrammatic working memory. The *If* part of the rule can also check for objects/relations in the diagram and the *then* part can make changes to the diagram. Fig 4 shows an example of the operator proposal and operator application rules in biSoar. Problem solving in this new bimodal Soar proceeds through the application of rules that act on and modify both the symbolic and diagrammatic sections of the working memory.

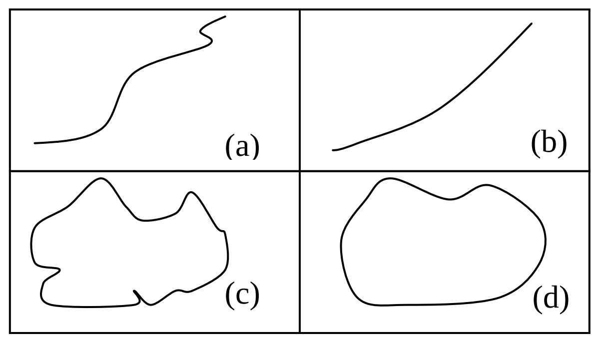


Fig 5: The effects of the visualize method.

Chunking in biSoar

The bimodal nature of LTM and WM in biSoar has consequences for chunking. Chunks (rules) that are learned in biSoar are bimodal. Like in chunking in Soar, only those representations (symbolic or diagrammatic) that were attended to during problem solving become part of the chunked rule. Diagrammatic chunking is implemented as a *visualize* method that is part of any routine that interacts with an object in the real world. The *visualize* method produces the equivalent of the product of attention on just aspects of the diagrammatic object. One way to think of *visualize* is that it is as if we are looking at the route at a very low resolution resulting in the loss of much of the finer details while still preserving the starting and ending points and the general curvature of the route. Fig 5(b) is the output of the visualize operator on the curve in 5(a) where the attention has been focused on just the betginning and end points. Fig 5(d) is similarly the result of visualize on Fig 5(c) where the attention has been focused on the braod shape, and on none of the details of the perimeter. The result of visualize does depend upon the requirements of the task because that determines the aspects to which attention was paid to in the diagram. But visualization in this manner is architectural because it happens irrespective of the task or the domain. For now, if attention is paid to a curve, it is considered to produce a straight line, but in general, it can thought of as producing a simple curve instead of a straight line.

Wayfinding

Wayfinding is a rich task domain to exercise a cognitive architecture intended to model spatial representation and reasoning. Human behavior in wayfinding exhibits a number of properties.

- Ability to recall the order of landmarks as well as relative orientation changes along the route is usually preserved.
- Simplification – routes recalled by subjects rarely preserve the exact curvature of pathways or their orientation to each other and to other landmarks. Curvature is usually straightened and actual angles are replaced by a small number of qualitative angular descriptions (left, slight right etc).
- Abstraction – multiple objects (whose details are irrelevant to the task) are usually abstracted to form a single object in recalled maps.
- Ability to hypothesize the destinations of untraveled routes, and the ability to generate novel routes.

The properties that are exhibited differ depending on the nature of wayfinding task as well as the source of the spatial information. For example, the ability to hypothesize destinations of untraveled routes is not required if the agent's source of knowledge is a map. On the other hand, this ability becomes important when the agent's knowledge about the world is created through the agent's interaction with the environment – that is, built up over time through the movement of the agent in the world.

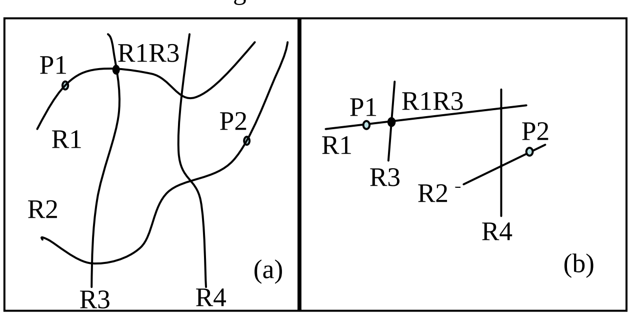


Fig 6: The map for the wayfinding task and the solution found by the model.

The term Cognitive Map is often used to refer to the various posited internal representations that constitute our knowledge of large-scale space. While there is disagreement as to the nature of these representations, there is a general consensus that the term map is misleading and that the representation is not a single, metrical, global map-like structure. Instead it is more likely a collection of fragments of knowledge of space, some metrical and some not, that are brought together to solve

0. Let the starting point be the current point.
1. Find the curve on which the current point lies.
2. Find the next point in every possible direction from the current point
3. For each possible next point, calculate the Euclidean distance between that point and the destination
4. Pick the point that is closest to the destination.
5. Make that the current point and repeat from 1.
6. Combine the curves generated so far to create the route from P1 to P2

Fig 7: wayfinding strategy used by the model

the spatial task. Cognitive Atlas (Hirtle 1998) and Cognitive Collage (Tversky 1993) are two of the suggested alternative terms that more correctly represent the nature of our representation of large scale space. In this task, we would like to explore how an architectural explanation of simplification can be given through the biSoar architecture. In particular, we explore whether the basic property of chunking (chunk only that to which attention was paid) is enough to explain the emergence of simplification in stored maps.

- The agent calls the DRS to find the routes on which P1 lies. (R1)
- The agent calls the DRS to find the next point in either direction on R1 from P1. (Start point of R1 and R1R3)
- Agent calls DRS to find the Euclidean distance between R1 and P2 and R1R3 and P2. (80 and 50)
- Agent picks R1R3 as the next point and calls for DRS to **copy** the route from P1 to R1R3. Finds the routes on which R1R2 lies. (R1 and R3)

Fig 8: one problem solving sequence from the task

The agent is given the task of finding a route from P1 to P2 in the map shown in Fig 6(a). The agent is asked to recall the route and the result is something like the map shown in Fig 6(b). Note that the recalled map has simplified paths. The wayfinding strategy used is shown in Fig 7. The problem solving sequence for one segment of the solution is shown in Fig 8. The critical step in the sequence is "copy". Copy contains the *visualize* method that produces the equivalent of the product of attention on the route from P1 to R1R3.

Discussion

The Spatial Semantic Hierarchy (SSH) provides a comprehensive theory of how an agent learns to build a representation of the environment, and how it uses this representation during problem solving. SSH 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). These representations are learned by the agent while navigating through an environment.

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 provides 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.

At this time, biSoar does not have an explanation for how representations are learned while the agent navigates the world. For now, biSoar's spatial knowledge comes from an external map. Also, certain issues, such as the control (sensor) level of the SSH, are outside the jurisdiction of the architecture. In biSoar, representations and processes at the control level would possibly be part of a perceptual/motor system with which biSoar communicates during navigation.

One of the advantages of developing models and theories within the framework of a general-purpose cognitive architecture is that it is sometimes possible to provide a single architectural explanation for phenomenon observed in different but related tasks. One such case is the phenomenon of simplification found in a multitude of spatial recall tasks. We have proposed biSoar (a bimodal version of the cognitive architecture Soar) and its diagrammatic chunking, an architectural learning mechanism similar to chunking in Soar, as one possible explanation for the simplification phenomenon. A model for the wayfinding task is used to describe how simplification emerges due to chunking. Due to limitations of space, we focused on the wayfinding example in the paper, but the same architectural feature of visualize as part of chunking can explain aspects of the phenomena related to recalling the directional relation between San-Diego and Reno. In particular, simplification affects what the agent learns about the shapes of California and Nevada and the ir orientation to each other. This later affects the agent when solving the problem of finding the relation between San-Diego and Reno. Our proposal has similarities with the proposal by Barkowsky for the use of visual representations in solving these kinds of problems (Barkowsky 2001).

Acknowledgements

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

References

Anderson, J. R., D. Bothell, et al. (2004). "An integrated theory of the mind." Psychological Review **111**(4): 1036-1060.

Barkowsky, T. (2001). Mental processing of geographic knowledge. Spatial Information Theory - Foundations of Geographic Information Science. D. R. Montello. Berlin, Springer**:** 371-386.

Chandrasekaran, B., Kurup, U., Banerjee, B., Josephson, J.R., Winkler, R. (2004). An Architecture for Problem Solving with Diagrams. In Proceedings of *Diagrammatic Representation and Inference*, Springer-Verlag: 151-165

Gillner, S. and H. A. Mallot (1998). "Navigation and Acquisition of Spatial Knowledge in a Virtual Maze." Journal of Cognitive Neuroscience **10**(4): 445-463.

Hirtle, S. (1998). The cognitive atlas: using GIS as a metaphor for memory. Spatial and temporal reasoning in geographic information systems. M. Egenhofer and R. Golledge, Oxford University Press**:** 267-276.

Kuipers, B. J. (2000). "The Spatial Semantic Hierarchy." Artificial Intelligence **119**: 191-233.

Kurup, U. and B. Chandrasekaran (2006). Multi-modal Cognitive Architectures: A Partial Solution to the Frame Problem. 28th Annual Conference of the Cognitive Science Society, Vancouver.

Laird, J. E., A. Newell, et al. (1987). "Soar: an architecture for general intelligence." Artificial Intelligence **33**(1): 1-64.

Lynch, K. (1960). "The Image of the City."

McNamara, T. P. (1986). "Mental representations of spatial relations." Cognitive Psychology **18**: 87-121.

Newell, A. (1990). "Unified theories of cognition." Harvard University Press.

Siegel, A. W. and S. White (1975). The development of spatial representations of large-scale environments. Advances in Child Development and Behavior. H. Reese. New York, Academic Press. **10:** 9-55.

Stevens, A. and P. Coupe (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. (1993). Cognitive maps, cognitive collages, and spatial mental model. Spatial information theory: Theoretical basis for GIS. U. Frank and I. Campari. Heidelberg-Berlin, Springer-Verlag: 14-24.