



Topics in Cognitive Science 3 (2011) 760–777

Copyright © 2011 Cognitive Science Society, Inc. All rights reserved.

ISSN: 1756-8757 print / 1756-8765 online

DOI: 10.1111/j.1756-8765.2011.01156.x

Augmenting Cognitive Architectures to Support Diagrammatic Imagination

Balakrishnan Chandrasekaran, Bonny Banerjee, Unmesh Kurup, Omkar Lele

Department of Computer Science and Engineering, The Ohio State University

Received 16 January 2010; received in revised form 4 February 2011; accepted 17 February 2011

Abstract

Diagrams are a form of spatial representation that supports reasoning and problem solving. Even when diagrams are external, not to mention when there are no external representations, problem solving often calls for internal representations, that is, representations in cognition, of diagrammatic elements and internal perceptions on them. General cognitive architectures—Soar and ACT-R, to name the most prominent—do not have representations and operations to support diagrammatic reasoning. In this article, we examine some requirements for such internal representations and processes in cognitive architectures. We discuss the degree to which DRS, our earlier proposal for such an internal representation for diagrams, meets these requirements. In DRS, the diagrams are not raw images, but a composition of objects that can be individuated and thus symbolized, while, unlike traditional symbols, the referent of the symbol is an object that retains its perceptual essence, namely, its spatiality. This duality provides a way to resolve what anti-imagists thought was a contradiction in mental imagery: the compositionality of mental images that seemed to be unique to symbol systems, and their support of a perceptual experience of images and some types of perception on them. We briefly review the use of DRS to augment Soar and ACT-R with a diagrammatic representation component. We identify issues for further research.

Keywords: Diagram; Diagrammatic representation; Diagrammatic reasoning; Mental imagery; Cognitive architecture; Visual memory

1. Spatial cognition, visual representations, and diagrams

The representational theory of mind proposes that our mental experiences and activities are underwritten by mental representations. Our concern in this article is with internal diagrammatic representations that we posit are created by agents in their minds when solving

Correspondence should be sent to Balakrishnan Chandrasekaran, Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210. E-mail: chandra@cse.ohio-state.edu

certain problems. We are concerned with what is needed for cognitive architectures to support such internal diagrammatic representations and operations on them.

This article is not about the importance of diagrams in reasoning—both cognitive scientists and AI researchers have written on the role external diagrams play in reasoning: for example, Gelernter (1963), Koedinger and Anderson (1990), and Lindsay (1998) in geometric proofs, and Larkin and Simon (1987) in problem solving in general. The article is also not about the differences as such between diagrams and other forms of representations such as text (for a discussion of this distinction, see, e.g., Stenning and Oberlander [1995] and Chandrasekaran [in press]). We are only concerned with internal representations, proposing that representations that can capture the functionality of diagrammatic imagery are a useful adjunct to those currently part of cognitive architectures. A stream of research emphasizes a role for perceptual, specifically visual, images in supporting certain kinds of problem solving. The status of mental images as representations has been a subject of philosophical and scientific debate; the following provide a critical summary of the issues: Pylyshyn (1973, 2002), Kosslyn and Pomerantz (1977), and Anderson (1978).

Visual experience in general supports seeing external objects as three-dimensional (3D) entities with color and texture. However, external diagrammatic representations are two-dimensional (2D) objects composed of points, curves, and regions. This distinction carries over to corresponding objects in imagination.

Visual imagery often involves rotation and translation in 3D (Shepard & Metzler, 1971). The corresponding imagistic operations on diagrams, on the other hand, are usually restricted to 2D. The representations and mechanisms that support such internal operations on diagrammatic imagery are unlikely to be dedicated to diagrams alone, but instead are instantiations of the corresponding ones for general visuospatial representations. Nevertheless, focus on internal diagrams can be useful because of the practical and theoretical importance of diagrams as representations. Diagrams—graphs, charts, maps, etc.—are ubiquitous in modern life, and predictive modeling of human performance in diagrammatic problem solving can be useful in the design of diagrammatic interfaces. Even when the underlying situation is visual, answering questions often involves diagrammatic abstractions. In answering the question, “Was John standing closer to Bill than to Stu at the party last night?” a person would typically construct a mental diagrammatic abstraction of points from his visual memory of the evening before. Diagrams are also interesting from the perspective of logic of representation, e.g., points and curves in diagrams are Euclidean abstractions, and the specificity of diagrams can be an asset as well as a liability in making inferences. We think that our proposal for how architectures can support diagrammatic imagination can be suggestive of solutions for the case of general visual imagination, but that is not our focus.

The current family of general cognitive architectures, of the sort exemplified by ACT-R (Anderson & Lebiere, 1998; Anderson et al., 2004) and Soar (Laird, Newell, & Rosenbloom, 1987), restrict their representations to predicate-symbolic representations for knowledge, goals, and cognitive states in general. Some of these architectures, for example Epic (Kieras & Meyer, 1997), ACT-R (Anderson et al., 2004), (which incorporates ideas originally proposed for Act-R/PM [Byrne & Anderson, 1998]) and CHREST+ (Lane, Cheng, & Gobet, 2000), do provide support for interfaces between perception and cognition

in using external representations. However, these architectures do not support visual imagery and operations thereon. CaMeRa (Tabachneck-Schijf, Leonardo, & Simon, 1997) is an exception to the exclusive focus on symbolic representations for problem solving. It provides support for an internal image in the form of a low-resolution visual bitmap as part of the subject's short-term memory that holds an image of the diagram, and support for some operations on this image. The work we report is part of a new direction of research in cognitive architectures with the goal of providing support for visual or diagrammatic imagery (Lathrop & Laird, 2007; Kurup & Chandrasekaran, 2007, 2009; Matessa, Archer, & Mui, 2007).

What functionalities are needed for a cognitive architecture to model internal diagrammatic representation as part of cognitive activity? We start by considering a set of examples of diagrammatic imagination and extract from such examples a set of constraints—design specifications—on any proposed representation. Admittedly, this approach of developing constraints from examples cannot provide a complete set of such constraints, but they do provide a good starting point. Then, we examine the degree to which DRS—for Diagrammatic Representation System—proposed in Chandrasekaran, Kurup, Banerjee, Josephson, and Winkler (2004)—satisfies the requirements for internal representation. We also discuss integration of DRS with symbolic cognitive architectures.

2. Constraints on internal diagrammatic representations

In what follows, we assume that internal diagrammatic representations exist as part of a cognitive architecture of the sort exemplified by ACT-R (Anderson & Lebiere, 1998; Anderson et al., 2004) and Soar (Laird et al., 1987). Their long-term memories store the agent's knowledge, and the control structures coordinate setting up and exploration of the relevant problem spaces by goal-relevant knowledge retrieved from long-term memory (LTM) and placed into working memory (WM). The diagrammatic representations that we posit are intended to augment the symbolic representations that exist in WM and LTM. Specifically, the imagistic representations exist in WM, and internal perceptions can be applied to them. We discuss some of the properties that these diagrammatic representations need to have to support diagrammatic imagination.

2.1. Object individuation and its implications (C1)

Internal representations are not just images, but consist of individuated addressable objects. Such a representation can be the result of perceiving an external diagram, imagining one, or a combination of elements from the external representation and imagination.

When there is an external representation, external perception performs figure-ground (re)organization and delivers to cognition a configuration of individuated spatial objects. To answer the question for Fig. 1, the agent's internal representation has to be organized as a pair of objects, each provided with its spatiality (location and shape) and independently manipulable. There is widespread agreement that one of the first tasks of external perception

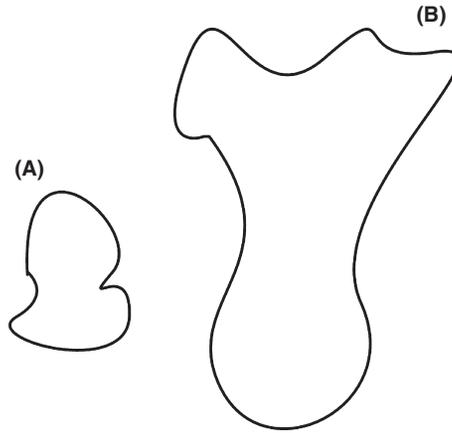


Fig. 1. A question requiring mental imagery operations: can region A fit into region B?

is to provide a figure-ground separation of an external representation (Kosslyn, 1989; Pinker, 1990) and to organize it as a configuration of spatial objects and groups of objects. This individuation property suggests that the underlying representation needs to support the representation of the diagram not as an array of pixels, but a collection of spatially specified objects.

Returning to Fig. 1, answering the question involves creating an internal representation composed of an element from the external representation (region B) and the result of mentally moving one of the objects (region A). In general, internal representations can be composed of elements from external representation, memory, and imagery operations on objects. The literature on graph comprehension—see Chandrasekaran and Lele (2010) for a review—provides empirical evidence for the use of imagery operations when using external representations. Cleveland and McGill (1984, 1985) make a distinction between elementary perception tasks, and those that require mental imagery operations. Simkin and Hastie (1987) additionally argued that even some of the elementary operations identified by Cleveland and McGill often required a sequence of visual operations, such as anchoring, scanning, projection, and superimposition—all operations on internal images. Gillan's research (Gillan, 2009; Gillan & Lewis, 1994) identifies a number of diagrammatic imagery operations involved in the use of external graphical representation: image move, image rotate, select anchor, image compare, and image difference. The comparison and difference operations involve moving one of the diagrammatic elements mentally to complete the task. Tricket and Trafton (2004) and Trafton and Tricket (2006), like Gillan, emphasize the importance of spatial transformations and imagination in graph comprehension. For example, using eye fixation and timing data, they demonstrated that subjects mentally extended a function line to predict y -values of the function for x -values that the graph did not cover.

Some of the steps in diagrammatic reasoning involve obtaining information from a diagram by performing relational perceptions, such as Inside in Fig. 1, and certain object property perceptions, such as noting that a curve is a straight line, or an angle is a

right-angle. These relational and property perceptions are performed on the object configuration in the internal representation, whatever the source of the configuration—external representation, memory, result of internal operations such as translation, or a composition of these. In Fig. 1, the relational perception operator $\text{Inside}(B, A)$ is applied not only to the external representation but also to the composition in which A has been translated. Similarly, whether the representation in Fig. 2 is available for inspection or is presented for a brief time and then taken away, answering the question, “Is B closer to A than to C?” is more or less equally easy, as long as the presentation of the diagram is for a long enough time, and the question is asked within a few seconds of the diagram being taken away. These examples provide support to the idea that relational perceptions and certain property perceptions are performed on the internal representation.

However, certain perceptions cannot be performed on internal representations; they require re-access to the external representation. It is sometimes the case that an external representation may be organized into more than one object configuration. Consider the well-known example in Fig. 3, which can be seen as a wine glass or two faces in profile. The former interpretation requires that the representation be seen as a region object (the cup), whereas the latter interpretation calls for the representation to be decomposed into two regions (face profiles) and two curves (top and bottom lines).

If the image in Fig. 3 is taken away after presenting it for a short time to a subject who sees this as a wine glass, it is very hard for the subject to see it as two faces. That is because, as we discussed earlier, seeing it as two faces requires a different figure-ground organization and decomposition into objects. The figure-ground (re)organization is a perceptual function that cannot be performed in the “mind’s eye,” but requires access to the external representation (Chambers & Reisberg, 1985). On the other hand, composition of objects, translation and some kinds of rotations of individuated objects, and applying relational perceptions to the objects is a task of cognition. This distinction provides a useful line of demarcation between external and internal images.

Complementary to obtaining information from the representation by perception is creation or modification of a diagram to satisfy symbolically expressed constraints, for example, “move region A to the right until all or most of it is inside region B,” for the task in Fig 1. Performing such actions require mediation by perception, for example, to judge whether an object has been translated far enough.

Human performance in the kind of perceptions relevant to diagrammatic reasoning, especially in graph comprehension tasks, has been studied by many researchers, with respect to external representations. Perceptions seem to be instantaneous (Cleveland & McGill, 1984, 1985; Simkin & Hastie, 1987) in some situations, whereas in others extended visual problem solving is called for, and this problem solving may display substantial individual variations.



Fig. 2. A configuration of points.

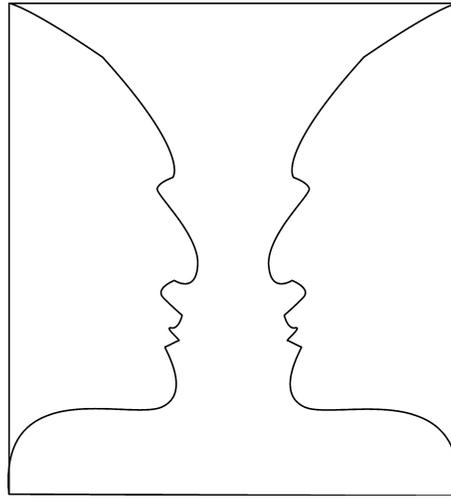


Fig. 3. An ambiguous figure.

Deciding which of two given curves is longer seems to be instantaneous if the curves are parallel straight lines close to each other, with one set of their end points suitably aligned. Variations on these requirements result in length comparisons calling for problem solving involving internal translations and perceptions. Currently, the field has empirical human performance data on only some comparison tasks (e.g., Simkin & Hastie, 1987; Gillan & Lewis, 1994; Gillan, 2009). Computational models that account for human performance over a range of such comparison tasks, including time taken and degree of errors, do not exist. Ullman's (1984) work on visual routines is suggestive of some ideas in this direction in that it identifies a primitive set of visual operators that are composed in a task-specific way to accomplish certain visual goals.

We can summarize the implications of the foregoing as follows. Any proposal for internal representation can assume that a process of external representation delivers a configuration of individuated diagrammatic objects, along with the spatiality of the objects. The internal perceptions need not include image reorganization as in the case of Fig. 3. However, the proposal should support creating configurations of objects from external representation and imagination, performing operations such as translation and rotation in the plane on selected objects, and applying a family of relational spatial perceptions on the configurations.

2.2. Points, curves, and regions as object types (C2)

In external diagrams, due to the requirements for external perception, all objects, including those intended to be points and curves, have non-zero spatial extent. That is, objects intended to be points and curves are regions, perhaps small circles representing points and strips denoting curves. Once their logical status as points and curves is recognized, the spatial properties added to help perception have no additional representational significance.

Similarly, once the objects in external representations corresponding to alphanumeric or other symbols (such as ‘‘A’’ and ‘‘B’’ in Fig. 1, or a Church icon at a location in a map) are recognized as the corresponding symbols or labels, the spatiality of the marks that make up the objects plays no further role. Internal diagrammatic representations do not need to incorporate this added spatial information, as long as they contain information about the intended object type (point, curve). For example in Fig. 1, the corresponding internal representation need only consist of two diagrammatic objects, one each for the two regions, with the abstract label A associated with the smaller region, and B with the larger one. The tokens A and B need not have corresponding diagrammatic objects.

2.3. Abstraction (C3)

Elements of external representation may be grouped and abstracted as objects, as may be useful for reasoning tasks. This abstraction may replace the original objects in the internal representation. In Fig. 4, the small circular regions on the left represent sensor fields whose range is a bit larger than the distance between the small circles. For deciding whether a vehicle in the center can get outside without being detected, the entire annular region can be abstracted as a sensor region, as shown on the right. The internal representation should enable the subject to change abstractions, seeing the set of regions as an annular region or as individual regions as the problem solving goals require. Abstraction is useful to control the complexity of internal representation, which is subject to WM limitations (Carpenter & Shah, 1998; Halford, Baker, McCredden, & Bain, 2005). Complex objects are represented hierarchically, with abstractions at the higher levels and details at the lower ones. Goal-oriented attention mechanisms enable focus on the relevant level of detail. Another example of abstraction being used to reduce complexity is when the degree of detail in an object, such as the wiggles in a complex curve, is not attended to unless specifically needed.

3. DRS, a proposed diagrammatic representation

The constraints we have outlined provided a starting point for our specification of DRS (Chandrasekaran et al., 2004). DRS is intended to be a domain-independent system for representing black-and-white line diagrams in general.

A diagram is a configuration of diagrammatic objects (see constraint C1), each of which is one of three types: point, curve, and region (C2). Point objects only have location (i.e., no

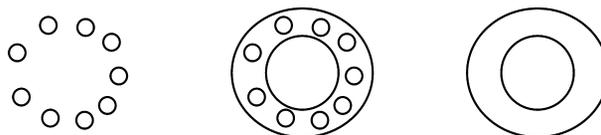


Fig. 4. For some problem-solving purposes, the cluster of small regions on the left may be abstracted into an annular region.

spatial extent), curve objects only have axial specification (i.e., do not have a thickness), and region objects have location and spatial extent. Associated with each object is the specification of the points in the 2D space that define the object, and additional features such as symbolic labels that are often attached to diagrammatic objects in physical diagrams.

As a data type, how the spatial specifications are represented is left to the implementation. It can be implemented extensionally, such as by specifying the intensity values of the elements in a 2D array. It can be an implicit description, such as algebraic expressions or equations that describe point, curve, or region objects. As the objects are typed, whether a closed-curve specification refers to a curve or a region is not in doubt. Fig. 5 is an example DRS for a simple diagram. A diagram in DRS is more than a raw image; it is the result of a figure-ground discrimination already made, in which the image array is interpreted as objects along with their spatial specifications (C1). Not all of what might be seen on a physical paper diagram will appear as spatial information in the DRS. The physical diagram may have a small circular region intended to represent a point, or a thin ribbon of a region to represent a curve, but the DRS version of a curve or point will simply be the intended point—the center of the circle, perhaps—or the intended curve—what will remain if the ribbon became infinitesimally thin—not the object as it appears on paper. Thus, points and curves in DRS are Euclidean.

Similarly, all aspects of the physical diagram whose sole purpose is to indicate something symbolic, but otherwise have no spatial significance, will be represented as symbolic annotations in DRS. Colors, icons, such as the picture of a church placed on a map to indicate the presence of church at that place, and hash marks in a region in military maps to indicate a no-go area, are examples. The corresponding DRS will not have colors, pictures of a church, or hash marks—the intended symbolic annotations would be attached to the corresponding objects.

DRS can be hierarchical. The gestalt region corresponding to the cluster in Fig. 4 might be represented as a region, with the small regions that produce the abstraction as subregions in it (C3). Seeing the cluster as a region would require appropriate gestalt perceptions.

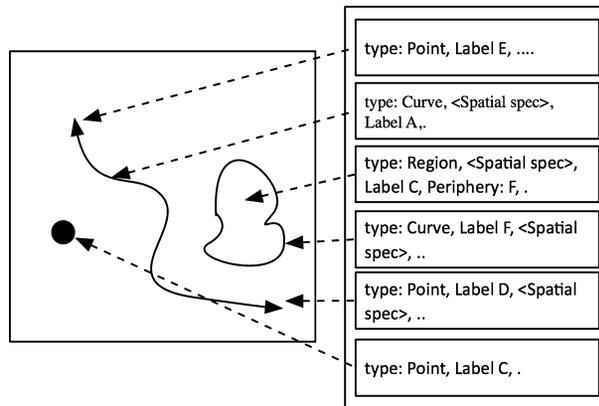


Fig. 5. A DRS for a simple diagram composed of a curve and a region. However, note that there are other objects: distinguished points such as end points, the closed curve defining the perimeter of the region, etc.

DRS is generic. Information that is domain-specific, for example, that a certain region is a “no-go region,” is simply treated as abstract symbolic information and incorporated as labels attached to the objects to be interpreted by the problem solver using domain-specific conventions. At the level of DRS, there is no recognition of a curve as a straight line or a region as a polygon. DRS simply records the objects that emerge after the figure-ground separation. We posit that seeing a curve as an object from the background precedes it being seen as a straight line, and seeing it as a straight line may not be relevant for some domains. Such characterizations can be attached to the objects as labels as soon as perception processes downstream recognize the properties. Abstract objects, such as the annular region in Fig. 4, are a special case. The initial DRS would only have the objects corresponding to the small circular objects. In many applications, it may be useful to have a certain set of gestalt abstractions, such as seeing a cluster as a region, invoked automatically, without a problem solver specifically asking for such abstractions. For example, if the DRS corresponding to the set of small regions in the left-most part of Fig. 4 is present, an initial set of perceptions that abstract the annular regions on the right may be automatically invoked, and a hierarchical DRS may be constructed, whose top levels correspond to the regions on the right of Fig. 4, and the lower levels to the smaller objects that constitute the annular region.

DRS representations can be constructed, as the needs of problem-solving dictate, as a composition of elements from external representation and memory (C1). The availability of a hierarchical DRS makes it possible for attention to shift from and to the abstraction and its components as needed. The change of focus of attention from an abstraction to the details is facilitated by the fact that a DRS representation is a hierarchical structure of individuated objects.

In keeping with the functional notion, a diagram in DRS contains the spatial specifications for a diagrammatic instance (not a class), and its spatial specifications must be complete, just as in the case of an external diagram, and unlike in the case of predicate-based representations that can provide partial specification of a situation or of classes of situations (such as for all). This does not mean that the agent is committed to all the spatial details in the DRS—it is the task of the problem solver to keep track of how the diagram represents, for example, which diagrammatic properties represent which domain properties. For example, a problem solver might treat a concrete diagrammatic instance of a right-angled triangle drawn on a piece of paper as really representing the class of such triangles, but that is a kind of symbolic annotation attached to the DRS of the concrete diagram.

3.1. DRS implementation

A set of requirements on representations, of the sort we compiled in the previous section on internal diagrams, cannot in principle be complete. Additional constraints, such as neural plausibility, processing time, etc., can be added. We have described DRS somewhat abstractly: The reader might note, for instance, that the representation of the spatiality is left to the choice of the implementer. What is part of the theory and what is to be left for implementation does not have a clear line—it is a matter of the theory maker’s interest at a point in time. One set of constraints we explicitly did not include is fidelity to timing and error

properties of human perception, as empirically available information is not yet as comprehensive as needed. How spatiality is represented is a key determinant of the computational behavior of perception algorithms. Instead of providing a specification that might later need to be changed to accord with empirical data on this score, our design approach is to specify it abstractly. Specific implementation may make additional commitments as needed for the purpose, and as more is known about the constraint, some implementations would accord with reality better than others.

DRS as described so far is minimalistic: We were only committed to supporting object individuation and related constraints discussed earlier. Implementing DRS as part of a cognitive architecture, as AI technology or for cognitive modeling, requires additional commitments, in particular for representing the spatiality of the objects. In our implementations, we explored two kinds of representations for spatiality. The first was in a purely algebraic framework (Banerjee & Chandrasekaran, 2010a): curves, as either as objects or as closed curves describing the peripheries of region objects, are specified as algebraic equations. In the second, the objects are represented in 2D arrays (Banerjee & Chandrasekaran, 2010b), similar to those used in the WM visual representations in CaMeRa (Tabachneck-Schijf et al., 1997) and Lathrop and Laird (2007). The two types of implementation make possible different algorithms for perception and diagram creation/modification with different efficiency properties. The array representations are posited by the above authors as capturing some of the properties of human internal perception.

In our current implementation, the frame in which internal images occur is square and the units are abstract rather than inches or centimeters. These choices were made not on the basis of any theory or empirical data: they seemed to be reasonable choices, but can be easily changed if required. In our implementations, the same shape in two different orientations would receive two different spatial extent specifications. Seeing them as the same shape in two different orientations is left to further operations such as mental rotation.

Now let us consider the implementation of perceptions and diagram creation and modification. For diagrams represented in DRS to be useful in problem solving, algorithms are needed to compute perceptions on objects in the diagram, that is, to compute specified spatial properties of diagrammatic objects and spatial relations between objects, and to create or modify diagrams in response to problem-solving goals. General frameworks for composing such algorithms have been developed by Banerjee (2007) for both implementations, array, and algebraic. These algorithms can also modify or create diagrammatic objects satisfying given constraints, and add them to DRS. The algorithms can detect emergent and vanishing objects as objects are added or removed from a diagram. Internal perceptions can be applied to a composition of diagrammatic objects from external representation, memory, and imagination. None of these algorithms is intended to simulate the corresponding algorithms in the human architecture, and hence they are not useful in predicting the timing and error properties of human performance.

A final issue we consider is the storage of diagrammatic elements in LTM. Imagination operations may require that a DRS in WM be augmented with diagrammatic objects that are in LTM. Conversely, learning by a mechanism such as Soar's chunking (Laird et al., 1987) may result in diagrammatic configurations being stored in LTM. In the current

implementation, individual objects and configurations are stored in LTM in a position-neutral way, so that when the location of the object/configuration is specified, the DRS in WM can be updated or constructed with the object/configuration in that location. By default, if no additional constraints are given, the object/configuration is placed in the center of the frame in WM during such construction. Otherwise, the various action routines have defaults for placing objects onto the frame, and these defaults can be overridden by additional specifications.

DRS has points of contact with other proposals to bring imagery inside cognition. The models in CaMeRa (Tabachneck-Schijf et al., 1997) and Lathrop and Laird (2007) have the individuation of objects and compositionality properties of DRS.

4. Integrating DRS in cognitive architectures

When diagrams are used in problem solving, only some of the steps involve the diagram; the others involve operations on symbolic representations, steps that are well-supported by existing cognitive architectures. As such, diagrammatic representation and operations on them have to be integrated into a general cognitive architecture. The principle that we use to perform such integration is that whenever there is a subgoal that requires access to the diagram, the diagrammatic representation and the associated operations are accessed, the perceptions or actions are applied so as to satisfy the subgoal, the diagram is updated as appropriate, and any relevant symbolic information is passed on to the symbolic component.

One way to integrate DRS with cognitive architectures is modular: The control component of the main architecture calls on a diagrammatic module to solve subproblems that require access to the diagram. The module has the diagram represented in DRS and comes with a set of perception and action operators. The main architecture knows what subgoals require the DRS component, and the module returns the relevant symbolic information to the main part. Matessa et al. (2007) is an example of such an approach, where ACT-R is augmented with a DRS-based diagrammatic module. A similar approach is taken by Lathrop and Laird (2007) in their integration of a visual simulation component with Soar. In these approaches, the core representations of ACT-R or Soar remain unchanged.

An alternative approach, one that may be called “integrated,” makes the diagrammatic representation and associated operations an integral part of the architecture. The biSoar effort (Kurup & Chandrasekaran, 2007), based on a theoretical stance about the multimodality of the cognitive state (Chandrasekaran, 2006), makes all cognitive state representations—in goals, WM states, the state descriptions in production rules—bimodal; for example, all states have, in addition to the traditional predicate-symbolic component, a diagrammatic component, represented in DRS, that depicts the visualizable aspects, if any, of the state. Just as symbolic operators are available to operate on the predicate-symbolic state, internal perception operators are available to solve relevant subgoals. Soar’s design is unchanged in all other respects. First, we will briefly describe biSoar’s design, specifically about how DRS is integrated. Later in this section, we will make additional comparisons between the two approaches.

Soar provides representations for short and long-term memory, mechanisms for interacting with the external world, a subgoaling 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 subgoals. The Soar architecture also provides a rule-based programming language that can be used to program the Soar agent. 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. Actions involve proposing and applying operators, which can be thought of as the next possible step to take in the problem-solving process. When the LHS of an LTM rule matches the contents of WM, the action part of the rule is instantiated appropriately and added as the current goal. We can call the contents of Soar’s WM and the operator, if any, that has been selected, the cognitive state of the Soar agent. If problem solving can be performed all within the symbolic framework and if LTM has sufficient relevant knowledge, this process will continue until the subgoals are eventually solved and consequently the original problem.

In the following description, we focus only on aspects of Soar that are relevant to the issues of integrating DRS and using it to do problem solving. For clarity, we use a notation that captures the spirit of Soar without following its syntax closely.

Soar’s representations are predicate-symbolic. Consider the scenario in Fig 6. Regions R1, R2, and R3 are impassable areas, while region S1 is a sensor area. Any object passing through a sensor area sets off an alarm. An object O1 is observed at location A at time t1 and another object O2 is observed at location B at time t2. The goal is to decide whether O1 and O2 are the same object or two different objects.

The representation in Soar’s WM corresponding to the initial state would consist of symbolic expressions such as ‘NoGoRegion(R1), NoGoRegion(R2), NoGoRegion(R3), SensorRegion(S1), Object(O1), Object(O2), At(O1,A,t1), At(O2,B,t2)’ (shown in the symbolic component in Fig. 6). Note that the objects are simply labeled and their types declared, but the representation doesn’t say anything about their spatiality. Several subgoals in this task require access to diagrams and application of perceptions or actions, and so some kind of

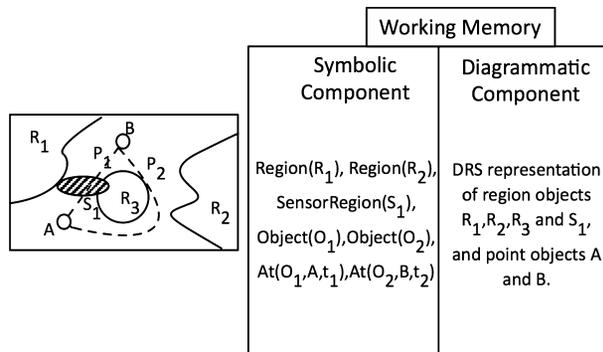


Fig. 6. The diagram on the left has four regions R1, R2, R3, and S1; two point objects A and B; and two curve objects P1 and P2. The WM of Soar would consist of just the symbolic component, whereas biSoar’s WM would also have the diagrammatic component, which has the diagrammatic elements represented in DRS.

diagrammatic ability would be needed; otherwise, Soar will stop without a solution. To be concrete, here are some of the initial steps in the problem-solving strategy: (a) to identify paths that the object at A might have taken to reach B; (b) see if any of the paths are short enough for O1 to have made it in the time available given its maximum speed; for each such path; (c) check if it crosses the sensor field; (d) if it does, check if there is information from the sensors about an object crossing; and (e) if there is not, see if the path could be modified so as not to cross the field, which would explain why there was no sensor report. At least for human problem solving, steps (a), (c), and (e) would require access to the diagram and application of perception. Steps (b) and (d) are symbolic processes: The former is numerical calculation, and the latter calls for database look-up. In the integrated approach of biSoar, the cognitive state is bimodal—biSoar's WM in Fig. 6 has both symbolic and diagrammatic parts. The reader should remember that the diagrammatic component is not simply a collection of symbols as it appears on the RHS of Fig. 6, but that the spatiality of the objects are all represented in DRS, so that functionally the diagrammatic component is a diagram. In general, the bimodal version of an LTM rule would have the form:

‘‘If $\langle S, D \rangle$, change state to $\langle S', D' \rangle$ ’’

where $\langle S, D \rangle$ specify the symbolic and diagrammatic components in the state specification. Let us consider subgoal (a), finding possible paths from A to B, avoiding no-go regions, FindPath(A, B, Avoid [R1, R2, R3]). The subgoal will be present in the symbolic part of WM, having been placed there at the previous step in biSoar's attempt to solve if the two sightings refer to the same or different objects. The LTM rule for that goal likely had on its RHS a set of subgoals including (a). Let us say a rule in LTM exists for such path-finding subgoals. As the subgoal in this case is symbolically represented, there is no need for the LTM rule to have a D component on the LHS. The rule's action part would propose an appropriate operator in the perception/action repertoire of the DRS, the action FindPath (A, B, Avoid [R1, R2, R3]). In the next cycle, this operator would be applied, and the result, Paths P1 and P2, will be added to the DRS. Soar provides standardized access for reading and writing structures to the input-output section of the symbolic component of the WM, making it easy to integrate DRS with Soar. It is useful to design the perception/action algorithms that make changes to DRS, so that they change the symbolic component of the WM to reflect the changes to DRS. For example, after FindPath succeeds, such a rule would add the expressions, Path(P1) and (P2), to the symbolic component of WM. The symbolic component now is aware that two paths exist, though for any information about their spatial extent, the DRS representation would need to be consulted. Similarly, when the subgoal (c) is set up, DRS would be tasked with the corresponding perception and the result, that the path does cross the sensor field, would be added to the symbolic component. Note that the fact that the path crosses the sensor region is implicit in the diagrammatic component, but it is not explicitly represented there. But once perception takes note of this fact, it becomes an explicitly symbolic assertion to be added to the symbolic component.

Whereas the foregoing captures the essence of how diagrams represented in DRS are used in biSoar, there are other aspects of biSoar that we do not discuss here but are of poten-

tial interest in the overall context of diagrammatic representations in cognitive architectures. The first is that biSoar extends chunking, Soar's learning mechanism, to the bimodal representation, so that rules with diagrammatic components can be learned and become part of LTM.

Second is the set of issues surrounding matching the diagrammatic components in WM. The LTM rule for the path-finding example only had the S component on the LHS, although the general form of the bimodal LTM rule has both S and D components on both sides. Shape-matching is not as straightforward as symbol structure matching, since, with a tight enough tolerance, no two shapes would match, and with a loose enough tolerance almost any two would. For this reason, in the examples we have considered, it has been more natural to let perception produce a symbolic state description, and then match it to rules in LTM whose descriptions are relevant for the problem-solving context.

Soar's use of a production-like LTM that is matched to the contents of WM can be considered as an attentional mechanism as it restricts the amount of knowledge brought to bear on a problem. This attentional mechanism applies in biSoar to the diagrammatic component as well.

4.1. Modeling with biSoar

Let us consider one example of cognitive modeling with biSoar. Stevens and Coupe (1978) reported on the errors a geographic recall task: 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. The most common reason for this error is that people simplify the regions in their mind, such as reducing regions to rectangles and ovals for a first approximation. Barkowsky (2001) provide the first computational model for this phenomenon incorporating the above insight. The general consensus is that the phenomenon is a result of simplification due to attentional constraints, resulting in the creation of a simplified diagram in memory. This (or part of it) is stored in LTM and any recollection (reconstruction) of this memory for the purposes of reasoning results in this error. Is this an architectural feature? If so, it should be near universal, but as it happens, while most subjects display this behavior, not all do. A model based on an architecture that distinguishes between invariant architectural features and individual-specific parameters can give a richer explanation. Such models, for example, can explore if an individual decision to pay more attention to the outlines of the states, or the individual having certain kinds of background knowledge would result in a different performance. We built a family of models using biSoar (Kurup & Chandrasekaran, 2007). In the first, limited attention mechanisms produce in LTM simple shapes for the two states and their relative positions. This plus knowledge that San Diego is in California and Reno is in Nevada, and default procedures to place them somewhere in the middle of the states, produce a DRS from which the agent concludes incorrectly that the former is to the west of the latter. However, the biSoar model enables us to explore the changes in behavior with increasing attention and due to additional knowledge.

A brief comparison of modular and integrated approaches will be useful. For many purposes related to diagrammatic reasoning, the two approaches are functionally quite similar.

As long as the diagram in the DRS form is available along with an associated repertoire of perception and creation/modification algorithms, the basic machinery for obtaining information from the diagrams and supporting imagination by creating diagrammatic elements is in place in both approaches. In the biSoar discussion, we pointed to the generalization of chunking to DRS elements that result in RHS of production rules acquiring diagrammatic elements. The modular approach, at least as so far implemented, does not have this property.

A potential attraction of the integrated approach over the modular one is that states on both the left and right hand sides of an LTM rule can have diagrammatic components, and the diagrammatic as well as the symbolic parts of the LHS of the rules can be used to match the corresponding components in WM, or in goal descriptions. In our discussion on biSoar, we considered the problems inherent in shape-matching in the case where the LHS of the LTM rules have DRS components. Retrieving LTM rules by matching on the symbolic component is much more common, although there may be specific circumstances in which such shape-matching might be useful.

5. Concluding remarks

Diagrams are, a restricted type of visual representation. They are especially common as representations to help in problem solving, and their relative simplicity makes it possible to focus on issues such as compositionality and mental operations, for example, translation and rotation, more easily than in the general case. Our hope is that this initial foray into the special case of diagrams is a useful first step in building computational theories for general visual imagery, and eventually multimodal cognition (Chandrasekaran, 2006). From an analysis of examples, we identified a set of constraints for diagrammatic mental images and operations on them. We discussed how a representational approach called DRS satisfies many of the constraints. The perception algorithms that we have developed so far deliver the functionality but are not psychologically realistic. That means they cannot as currently implemented capture the timing and error rates associated with the corresponding human perceptions. We also briefly reviewed the ways in which DRS and the associated operators have been integrated within existing symbolic architectures such as Soar and ACT-R.

Perceptual imagination requires what anti-imagists thought was a contradiction: the ability of mental images to have the compositionality properties that seemed to be unique to symbol systems, and to support a perceptual experience of images and deployment of some types of perception on them. DRS, though restricted to a subtype of visual representation, shows how the apparent contradiction can be resolved: Mental images are not raw images, but a composition of objects, which can be individuated and thus be symbolized, while, unlike traditional symbols, the referent of the symbol is an object that retains its perceptual essence, namely, its spatiality. Mental images as well as the result of perception on external representations both give rise to the same perceptual experience because the locus of both of them is cognition.

We agree with Anderson (1978) that it is really hard to discriminate between representations in the mind, and that the highly abstract nature of the concept of propositions that are often contrasted with pictorial representations leaves open the possibility that pictorial representations can themselves be included in the propositional category. The DRS proposal is not a claim about what the “real” representation is inside the head, but it is instead motivated by the need to provide a certain set of functionalities. In this, it is similar to the production rule formalism and attribute-value representations that cognitive architectures support. They are a modeling language, as is DRS. In fact, it is precisely the desire to avoid over-commitment to representational details that motivated us to keep the specification of spatiality in DRS objects to be abstract, so that different formalisms can be explored.

The ideas underlying DRS have some points of contact with the proposal by Barsalou (1999) on Perceptual Symbol Systems, in which “subsets of perceptual states in sensory-motor systems are extracted and stored in LTM to function as symbols.” These symbols can be retrieved for later simulation and composed and abstracted to produce complex mental representations. His project is more radical—to remove altogether the need for a modal representations (what we have called predicate-symbolic representations in this article), but we don’t need to embrace that goal to see the commonalities with the idea of composable perceptual symbols providing a basis for some forms of reasoning, especially those involving simulation. Similarly, the proposal by Damasio (1994) that what brains encode are images of experiences in various modalities and mental activity is largely operations on these images has echoes in the DRS proposal for the limited domain of diagrammatic visual images.

Acknowledgments

This research was supported by the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009. The conclusions of this research do not necessarily represent the views of the sponsors. We wish to thank the referees and Glenn Gunzelmann for suggestions that helped improve the article.

References

- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85, 249–277.
- 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.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Banerjee, B. (2007). *Spatial problem solving for diagrammatic reasoning*. Ph.D. dissertation, Department of Computer Science & Engineering, The Ohio State University, Columbus.

- Banerjee, B., & Chandrasekaran, B. (2010a). A constraint satisfaction framework for executing perceptions and actions in diagrammatic reasoning. *Journal of AI Research*, 39, 373–427.
- Banerjee, B., & Chandrasekaran, B. (2010b). A spatial search framework for executing perceptions and actions in diagrammatic reasoning. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 144–159). Lecture Notes in Artificial Intelligence 6170, Heidelberg: Springer.
- 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.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.
- Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 167–200). Mahwah, NJ: Lawrence Erlbaum.
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4, 75–100.
- Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, 11, 317–328.
- Chandrasekaran, B. (2006). Multimodal cognitive architecture: making perception more central to intelligent behavior. *Proceedings of the AAAI National Conference on Artificial Intelligence*, 1508–1512.
- Chandrasekaran, B. (in press). When is a bunch of marks on paper a diagram? Diagrams as homomorphic representations. *Semiotica*.
- Chandrasekaran, B., Kurup, U., Banerjee, B., Josephson, J. R., & Winkler, R. (2004). An architecture for problem solving with diagrams. In A. Blackwell, K. Marriott, & A. Shimojima (Eds.), *Diagrammatic representation and inference* (pp. 151–165). Berlin: Springer-Verlag.
- Chandrasekaran, B., & Lele, O. (2010). Mapping descriptive models of graph comprehension into requirements for a computational architecture: Need for supporting imagery operations. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 235–242). Lecture Notes in Artificial Intelligence 6170, Heidelberg: Springer.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.
- Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, New Series*, 229(4716), 828–833.
- Damasio, A. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Avon.
- Gelernter, H. (1963). Realization of a geometry-theorem proving machine. In E. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 134–152). New York: McGraw-Hill.
- Gillan, D. J. (2009). A componential model of human interaction with graphs: VII. A review of the Mixed Arithmetic-Perceptual model. In *Proceedings of the Human Factors and Ergonomics Society 52th Annual Meeting* (pp. 829–833). Santa Monica, CA: HFES.
- Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs. I. Linear regression modeling. *Human Factors*, 36, 419–440.
- Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, 16, 70–76.
- Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 4(12), 391–438.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14, 511–550.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225.
- Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9, 52–76.
- Kurup, U., & Chandrasekaran, B. (2007). Modeling memories of large-scale space using a bimodal cognitive architecture. In R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 267–272). Ann Arbor: University of Michigan.

- Kurup, U., & Chandrasekaran, B. (2009). A cognitive map for an artificial agent. *Proceedings of the Second Conference on Artificial General Intelligence, AGI 2009, Arlington, Virginia*. Amsterdam-Paris: Atlantis Press, ISBN: 978-90-78677-24-6, 85-90.
- Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1), 1-64.
- Lane, P. C. R., Cheng, P. C.-H., & Gobet, F. (2000). CHREST+: Investigating how humans learn to solve problems using diagrams. *Artificial Intelligence and the Simulation of Behaviour Quarterly*, 103, 24-30.
- Larkin, J. H., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.
- Lathrop, S. D., & Laird, J. E. (2007). Towards incorporating visual imagery into a cognitive architecture. In R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 25-30). Ann Arbor: University of Michigan.
- Lindsay, R. K. (1998). Using diagrams to understand geometry. *Computational Intelligence*, 14, 238-272.
- Matessa, M., Archer, R., & Mui, R. (2007). Dynamic spatial reasoning capability in a graphical interface evaluation tool. In R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 55-59). Ann Arbor: University of Michigan.
- Pinker, S. (1990). A theory of graph comprehension. In R. Feedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73-126). Marwah, NJ: Erlbaum Hillsdale.
- Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychology Bulletin*, 80, 1-24.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Science*, 25, 157-238.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 171(972), 701-703.
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398), 454-465.
- Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, 19(1), 97-140.
- Stevens, A., & Coupe, H. P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10(4), 526-550.
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science*, 21(3), 305-350.
- Trafton, J. G., & Trickett, S. B. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, & N. Swoboda (Eds.), *Diagrammatic representation and inference* (pp. 286-300). Heidelberg: Springer-Verlag.
- Trickett, S. B., & Trafton, J. G. (2004). Spatial transformations in graph comprehension. In A. F. Blackwell, K. Marriott, & A. Shimojima (Eds.), *Diagrammatic representation and inference* (pp. 372-375). Heidelberg: Springer-Verlag.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 97-159.