Towards a Computational Model of Image Schema Theory

Wei Mu

Abstract

Embodied rationality has received increasing attention in artificial intelligence since the 1990s. One approach to embodied rationality is based on a concept from cognitive linguistics, image schemas. The theory of image schemas explores the ways that meaning, understanding, and rationality arise from and are conditioned by the patterns of our bodily experience. Image schemas provide a bridge between cognitive and bodily structures. We believe the way that humans reason intelligently, based on their bodily experience, can give us insight into how to design a computational model to achieve similar intelligent reasoning that is interpretable in human terms. This paper gives a survey of the existing research in the areas of image schema and mental imagery, focusing on the requirements for the representation. Extending previous efforts of computational models of image schemas, we have built a prototype that combines both depictive and propositional representations to capture some properties of image schema structures. We describe preliminary work with the prototype to explore its application to problem-solving.

1 Introduction

We, human beings, are "rational animals". This fact implies that our rationality is embodied. Bodily experience influences what and how things can be meaningful for us, the way we comprehend and reason about the world, and the actions we take. Embodied rationality has received increasing attention in artificial intelligence since the 1990s, in theories of agency and robotics. Mark Johnson's work, in The Body in the Mind, gives an overview of this area from a philosophical perspective that motivates the research presented in this paper.

To explain how embodied experience functions in our life, two types of imaginative structures are presented: image schema and metaphorical projection. An image schema is *a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience* [Johnson 1987, p. xiv]. Thousands of times each day we feel, see, or manipulate different kinds of image schemas, like containment (such as houses, cars, and boxes); like force (which causes a door to stay open or results from contact with others); like up-down orientation (which arises because we live in a gravitational field).

Generally speaking, image schema structures are more like a "plates" than the "food" inside. These structures allow us to store and categorize our basic physical and sensory experiences in different "meaningful" ways[1]. But the second imaginative structure, metaphorical projection, allows us to project patterns from one domain of experience into the structure of a different domain [Lakoff and Johnson 1980]. Here, *metaphor is not only a linguistic mode of expression; rather, it is one of the chief cognitive structures by which we are able to have coherent, ordered experiences that we can reason about and make sense of* [Johnson 1987, p. xv]. There are tons of examples can demonstrate the importance of metaphorical projection in helping people to connect knowledge from different domains. "More is up" is the most typical one. "More is up" implies that we understand quantity in terms of a verticality schema in our everyday experience. Examples like "Prices keep going up", "The number of books published each year keeps rising", "His gross earnings fell", and "Turn down the heat" all suggest that we understand MORE as being oriented UPWARD. This phenomenon isn't an accident. It's because we are living in a gravitational field, you will see the level go up when you add liquid into a container or put more objects to a pile. When the two schemas appear together in many cases, MORE and UP are therefore correlated in our experience that provides a physical basis for our abstract understanding of quantity.

Theories of image schemas and metaphorical projections allow us to understand more about meanings and rationalities by connecting them with our bodily experience. Furthermore, they provide a foundation for one of the most important reasoning capabilities of human beings, mental imagery, which is highly interesting for neuroscientists and psychologists [Kosslyn 1996].

---

[1] we discuss "meaning" in more detail in the next section

This view of rationality has a different flavor from much work in AI on abstract problem solving. We didn't start our research from one specific problem. Instead, we start our research from the fundamental theory of rationality, and try to implement the principle of this theory into a particular domain. The reason for the survey of these areas is that we believe the way that humans reason intelligently, based on their bodily experience, can give us insight into how to design a computational model to achieve similar intelligent reasoning that is interpretable in human terms. Therefore, a review of some current work in building a computational model using image schema theory is also presented.

This paper is organized in the following order. We first give an overview of a general theory of image schemas developed in cognitive linguistics. We then describe independent work on reasoning based on schematic representations, mainly in the area of mental simulation of physical processes. We describe how some of these concepts have been put together in preliminary work on computational models of image schemas. Finally, we present a representation and prototype implementation that shows some promise for being able to capture aspects of reasoning based on image schemas and to some extent metaphorical projection.

2 Image schemas as a foundation for rationality

2.1 Meanings are embodied

The research of artificial intelligence has been developing for more than half a century, and today's intelligent agents can exceed human performance in a lot of aspects, like "Deep Blue" has defeated the world chess champion Garry Kasparov. But there's always a question when people talk about the computer intelligence: "Although it can do this, does it really understand the meaning of what it's doing?" One of the famous questions is the "Chinese room" argument by John Searle in the early eighties. The core of Searle's argument is the distinction between syntax and semantics. That is, the room's behavior can be described as following syntactical rules. But in Searle's account it does not know the meaning of what it has done; in another words, it has no semantic content.

Mark Johnson introduces the theory of image schemas by first examining the question, “How can anything be meaningful to a person?” In a traditional sense, linguistic meaning, which assumes that only words and sentences have meanings and all of these meanings must be propositional, is the primary in the philosophy of language and linguistics. Objectivism treats all meaning as conceptually and propositionally expressible in literal terms that can correspond to objective aspects of reality. Johnson argues this is not enough because, 1, Meanings in natural language come with figurative and multivalent patterns that cannot be reduced to literal concepts and propositions; 2, the patterns and their connections are embodied and cannot be reduced to literal concepts and propositions. [Johnson 1987]

For example, one of the most important concepts we recognized from infancy is “force”. Thousands of times we move our body, manipulate other objects, and encounter obstacles; also we feel force exerted on us. From those ceaseless events, we develop repeatable patterns that we can recognize and carry out for interacting forcefully with our environment. These patterns are embodied and give coherent, meaningful structure to our physical experience at a pre-conceptual level, though we are eventually taught names for at least some of these patterns, and can discuss them in the abstract. But the meaning of such a pattern the meaning it identifies goes deeper than our conceptual and propositional understanding.

2.2 Linguistic interpretation of image schema

Inspired by linguistic and philosophical evidence, Mark Johnson and George Lakoff first developed a theory of image schema in late eighties. They both gave plenty of examples in linguistic expressions that evidence dynamic patterns of recurrent bodily experience. The notion of “schema” was based on Immanuel Kant’s work. For Kant, schemas are structures of the imagination; that is, schemas are fixed templates superimposed onto perceptions and conceptions to render meaningful representations [Oakley In press]. In Cognitive Linguistics, the term “image” implicates perception in all acts of conceptualization. Concepts develop from representation of a perceptual conglomeration of multiple senses, including visual, auditory, haptic, motoric, olfactory, and gustatory experiences. According to Johnson, image schemas are neither propositional, nor rich concrete images, nor mental pictures. They are structures that organize our mental representations at a level more general and abstract.

Figure 1: PATH schema

Typically, an image schema will have parts and relations. The parts represent a set of entities, and the relations represent the connection among different parts, like causal relations, temporal sequences, part-whole patterns, etc. Figure 1 can be thought of as a PATH schema. It contains three parts, a source point A, a terminal point B, and a trace between them. There also is a relation in this schema. The basic meaning of this schema can be understood as moving from A to B in a physical environment. Here, the schema represents a physical connection. But this schema also can be used to understand the sentence like "the melting of ice into water", in which the schema is applied at a more general and abstract level. But the relations between A and B stay the same.

In another example, Johnson examines six types of "out" in English:

1. Mary got out of the car.
2. Whenever I'm in trouble, she always bails me out.
3. I don't want to leave any relevant data out of my argument.
4. Don't you dare back out of our agreement.
5. Honda just put out its 1986 models.
6. We kicked him out of the club.

He argues that all of them relate to the IN-OUT schema shown in figure 2. He identifies the two parts as "landmark" (LM) and "trajectory" (TR). This schema clearly represents the visual situation of the first sentence. Here the circle represents the car, and Mary moves along the arrow out of the car. Since the car can't be circular, Mary can't move along a straight line in leaving the car, so this schema gives us only one idealized image. In the rest of the sentences, Johnson explains the pervasive act of

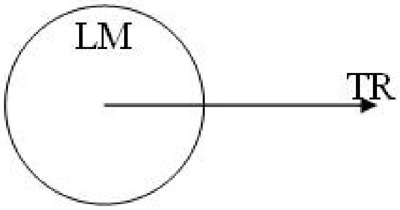


Figure 2: IN-OUT schema

metaphorically extending a schema from the physical to the nonphysical.

Lakoff also analyzes the linguistic senses of the word "over", like "the fly is over my head." and "I turned the log over." He uses an ABOVE image schema, where a small trajectory passes over a landmark, to explain the sentences. In their new model, language makes much more use of the brain's processes of spatial, visual and mental imagery than previously thought.

From the above examples, we can see that image schemas are dynamic patterns rather than fixed and static images, as their visual diagrams represent them. They are dynamic in two important respects: (1) schemas are structures of an activity by which we organize our experience in ways that we can comprehend; (2) unlike templates, schemas are flexible in that they can take on any number of specific instantiations in varying contexts.

2.3 Image schema: the body within the brain

Only having the structure is not enough for image schemas to support rational inference. We can perform mental operations on image schemas that are analogs of spatial operations. For example, we can rotate mental image both in 2D and 3D situation, and it seems we can do that at a fixed rate of approximately 60 degree per second, as studied by Shepard and Metzler.

Lakoff also indicates four primary schematic operations, called image-schema transformations,

- *Path-focus to end-point focus:* imagine the path of a moving object and then focus attention on the point where it comes to rest or where it will come to rest.

- *Mutiplex to mass:* imagine a cluster of objects. Now imagine moving away from the cluster until the individual objects start to appear as a homogenous mass.
- *Trajectory:* mentally traverse the path of a continuously moving object.
- *Superimposition:* imagine a large sphere and a small cube. Now, increase the size of the cube until the sphere can fit inside it. Now reduce the size of the cube until it fits back inside the sphere.

Along with other research in the neurosciences, people have found more and more evidence indicates activation patterns like image schemas are exist in both animal and human brains [Rohrer 2005; Johnson and Rohrer 2006]. Rizzolatti showed macaque monkeys visual imagery of another monkey grasping a banana with their hands, they were able to record activity from "mirror" neurons in the same areas of secondary somatomotor cortex that would be implicated if the monkey himself were performing the particular grasping action. Similar human experiments in which participants watched a video clip of another person performing an action showed increased activation in the corresponding area in human brain which is known to map human hand and arm grasping motions.

Also the research from developmental psychology suggests that infants come into the world with capacities for experiencing image schema structures. Stern studied the types of experiential structure that infants are able to detect, like cross-modal perception and vitality affect contours [Johnson and Rohrer 2006]. "Cross-modal perception" implies that even an infant can take information received in one sensory modality and somehow translate it into another sensory modality. Vitality affecting contours is the second type Stern studied. This experiment suggests that humans respond to the emotional satisfaction that comes from pattern completion, and witnessing even just a portion of the pattern is enough to affect contours in motion. For example, the infant just needs to see us begin to reach for the bottle, and she already begins to quiet downthe grasping image schema does not even need to be completely realized in time before the infant recognizes the action.

To summarize, image schemas can be characterized more formally as [Johnson and Rohrer 2006],

- recurrent patterns of bodily experience,

- that are “image”-like in that they preserve the topological structure of the perceptual experience,
- operating dynamically in and across time,
- realized as activation patterns (or “contours”) in and between topologic neural maps,
- they are structures that link sensorimotor experience to conceptualization and language, and
- support “normal” pattern completions that can serve as a basis for inference.

2.4 Metaphorical interpretation

Even though image schemas are generated from physical body experience, they can be applied to many other aspects of our lives, and they can help us to comprehend and make inference on other experience. Metaphorical interpretation[2] is the bridge by which we can understand and structure one domain of experience in terms of another domain of a different kind.

By the idea of metaphor, we know that metaphor structure should have two main roles: a source domain — from which we draw metaphorical expressions; and a target domain — the one we try to understand. A mapping is the systematic set of correspondences that exist between constituent elements of the source and the target domain. To know a conceptual metaphor is to know the set of mappings that applies to a given source-target pairing.

Let’s consider an example to explain how conceptual metaphor is so important abstract conceptualization and reasoning. The container schema is one of the most common conceptual abstractions from our bodily experience. A containment relation has several properties. First, an entity is either inside or outside the container but not both at same time. Second, if object A is contained by object B, and object B is contained by object C, than object A must be in object C. Now, let’s consider another concept — category, and the common metaphor, “Categories are containers”. In this metaphor, we can understand a conceptual category by an abstract container for physical and abstract entities. Therefore, we can say that “humans” is either contained in category “animals” or outside category “animals”. If “humans”

---

[2] In later work, Lakoff and Johnson call this conceptual metaphor.

is contained by “animals”, then because “animals” is contained in “living things”, “humans” are also “living things”. Johnson gives another example, “sexual appearance is a physical force” in his book. In this example, people project meaningful structure from our experience of forceful interactions of physical objects and events onto the domain of sexual experience.

By using conceptual metaphor, Lakoff, Johnson, and others are developing a rapidly growing body of metaphor analyses of key concepts in nearly every conceivable intellectual field and discipline, including the physical and biological sciences, economics, morality, politics, ethics, philosophy, anthropology, psychology, religion and more. For example, Lakoff and Núñez [2000] have carried out extensive analyses of the fundamental metaphorical concepts that underlie mathematics, from simple models of addition all the way up to concepts of the Cartesian plane, infinity, and differential equations.

However, metaphorical projection typically goes from the more concrete to the more abstract, and not the other way around. Accordingly, abstract concepts are understood in terms of prototype concrete processes. Consider the sentence “We have a long way to go before my work is finished.” Clearly, this sentence use physical PATH to interpret a social concept. “Move is up”, “Time is money”, and “problems are puzzles” are other good examples, which we use concrete concept to understand an abstract one.

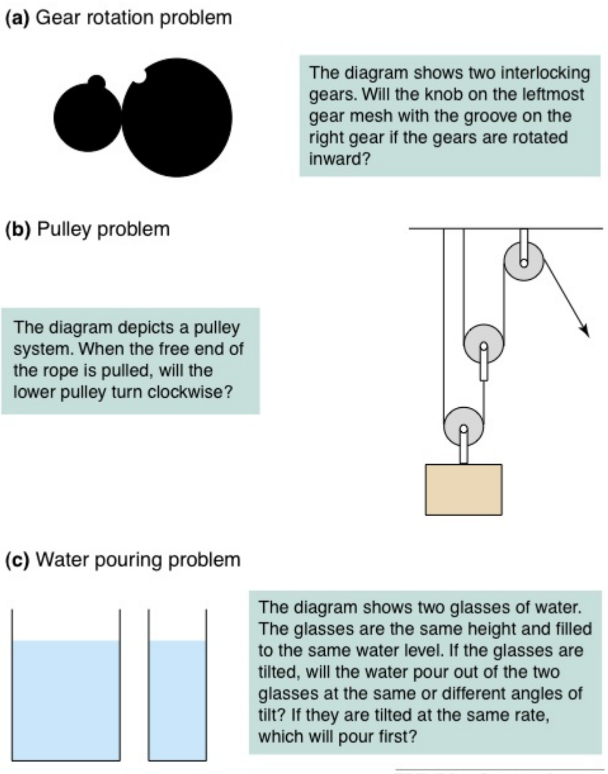
3 The need for a richer representation

3.1 Simulations in mental imagery

The nature of image schema theory is to provide a mechanism for people connecting the bodily experience with conceptual reasoning. It mainly focuses on where the materials of mental activities come from and how they are interpreted. Research in this area doesnt study too deeply the mechanism of how the brain actually does the reasoning using those experience or concept. It’s like giving you parts of a computer but not telling you how to connect each part to build an actual computer. However, another research area, mental imagery, does fill the gap.

As a collection of psychological phenomena, mental imagery has long held the interest of the scientific community. Mental imagery can be defined as “the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event [Bertel et al. 2006]". As early a scientist as Plato has tried to explain imagery with his famous wax tablet metaphor. Due to the lack of a method to actually measure or analyze imagery in human brain, mental imagery hasnt been the central concern of psychologist until recently.

Mary Hegarty has studied mechanical reasoning by mental simulation. In her paper [Hegarty 2004], Hegarty claims that a mental model (or situation model) is a representation that is isomorphic to the physical situation that it represents, and the inference processes simulate the physical processes being reasoned about. In other words, people will consciously simulate what will happen when they are solving mechanical problems. Hegarty uses the examples in figure 3 to illustrate this phenomenon.



The diagram shows two interlocking gears. Will the knob on the leftmost gear mesh with the groove on the right gear if the gears are rotated inward?

The diagram depicts a pulley system. When the free end of the rope is pulled, will the lower pulley turn clockwise?

The diagram shows two glasses of water. The glasses are the same height and filled to the same water level. If the glasses are tilted, will the water pour out of the two glasses at the same or different angles of tilt? If they are tilted at the same rate, which will pour first?

TRENDS in Cognitive Sciences

Figure 3: Examples of mechanical reasoning problems

In her analysis, Hegarty also emphasizes spatial ability more than verbal ability for internal representation. Here, spatial ability is characterized as the ability to construct, maintain, and transform spatial representations accurately.

Hegarty points out three properties in mental simulation,

1. People mentally simulate the behavior of complex mechanical systems piecemeal rather than holistically.
2. Mental simulation is not based purely on visual information, but also incorporates information about invisible entities and properties, such as force and density.
3. Mental simulation may involve motor representations as well as visual representations.

3.2 The debates of mental imagery

In research on mental imagery, the nature of mental imagery representations has been debated for many years. The center of this debate is whether visual mental images rely on depictive representations, or whether they are purely propositional representations. This is critical because when researchers began to think about how one could build a computational model to mimic imagery, one must specify a representation with particular properties.

Stephen Kosslyn and Zenon Pylyshyn are two of the most important figures in this area. They have done lot of research in mental imagery and hold opposite views in this debate. Kosslyn designed a series of experiments [Kosslyn 1996] that focused on the so called "privileged properties of depictive representations", which are not shared by propositional representations.

"Mental Scanning" (see figure 4) is one of them. It focused on the fact that depictive representations use "functional space" to represent actual space. In this experiment, participants learned a map, such as the following one. They were then asked to imagine the map, fix their attention on a given landmark, and indicate when they could "see" a second named landmark in their image. A linear relationship was observed between reaction time and distance between the two landmarks. In other words, the greater the distance between the two landmarks, the longer the reaction time a participant

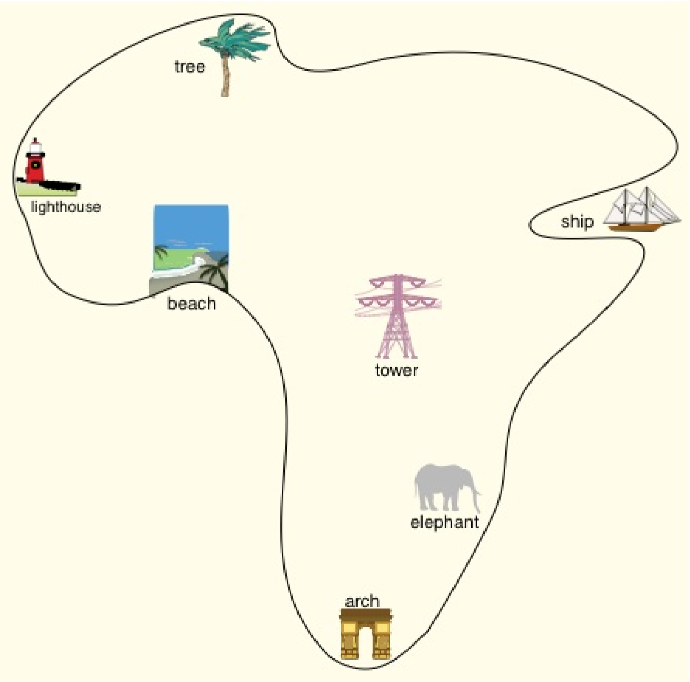


Figure 4: The experiment of mental scanning

needed.

One explanation is that mental imagery does have depictive representations involved, so that people need to use more time to "scan" the space between two landmarks. However, propositionalists also suggested that in purely propositional representation, both the landmark and the open space are represented as propositions in a "mental list", and the result only reflect the time to work down mental lists of propositions. Kosslyn and his colleagues conducted many experiments to rule out the later explanation. For example, they varied the distance and the number of objects that participants scanned over in an image, and showed that participants required more time to scan greater distances per second.

In the so-called the second stage of imagery debates, Pylyshyn suggested that there are potential methodological problems involved in experiments on mental imagery. For example, one problem is that the very act of asking subjects to use imagery may lead them to try to mimic what they would do in the corresponding perceptual situation. Another concern is the experimenters may have unconsciously led the subjects to produce the expected result. Although different experiments have been conducted to rule out those concerns, many researchers found it difficult to rule out the possibility that unconscious motives and beliefs govern subjects behavior in imagery experiments.

In the latest progress in the debates, many results from neuroscience have been found to support the idea that to have a mental image is to project two-dimensional moving pictures onto the surface of your visual cortex [Kosslyn 1996]. This idea has been fostered by the following findings: (1) When a visual pattern is presented to the eye, a homeomorphic (continuously deformed) mapping of retinal activity occurs in visual cortex. (2) Although it remains controversial, it has also been reported that there is increased activity in retinotopically-organized areas of visual cortex during mental imagery.

Pylyshyn [2003] argued that these findings tell us nothing about the form of the representation, because imagery and vision might involve the very same form of representation without it being pictorial in either case. Furthermore, Pylyshyn was concerned that there may be an infinite regress in pictorial mental imagery: if something is interpreting the image, what inside that something must be interpreting its interpretation of the image, and so on?

Although the debate about the nature of mental imagery is still not resolved yet, it doesn't affect the idea that pictorial depiction is involved in mental imagery. It may be at some layer in the middle, or maybe pictorial depiction is simply part of the nature of mental imagery. The important thing about this debate is that it will help us to realize what kinds of properties are involved in depictive representations but not in propositional representations. By analogy, when you program an algorithm on computer, you can ignore the nature of a computer is just some electron activities, but focus on a higher and more abstract level. That is, we can ignore the underlying disagreement about whether a depictive representation is based on a propositional representations or is primitive; the evidence is compelling that *depictive representations play an important role in mental imagery.*

3.3 Requirements for mental representation

Bertel et al. [2006] summed up some basic work on mental imagery and, more importantly, tried to formulate the requirements that should be reflected in computational mental imagery modeling.

The first thing Bertel discussed is whether mental imagery is like pictures or like propositions. In the early years, propositional structures have been used for almost all cases. However, there seems to be more than the visual part to a mental image; it is also likely that non-visual mental reasoning operates on more than just propositions. More recent analyses suggest that neither format may be adequate. Bertel said that mental images should be conceived of as hybrid, exhibiting both visual and propositional traits.

Secondly, mental images are constructed from pieces of knowledge that are retrieved from memory. These knowledge fragments are either elementary or they may be further structured like memory chunks and consist of an aggregation of a (small) number of elementary knowledge fragments.

Next, Bertel pointed out that knowledge from memory is often underconstrained with respect to what is needed for image construction: it can be incomplete, scarce, or “lean”. Therefore, during the construction, mental processes need to dynamically add knowledge fragments to increase specificity. Furthermore, the knowledge also could be distorted or partly conflicting. Some processes that resolve conflicts at least locally are needed in order to achieve the necessary level of knowledge integrity.

From research in mental imagery, it’s not hard to see that spatial reasoning is one of the most important capabilities of human intelligence. We take the hypothesis that a propositional representation, without processing specific to image schema properties, is insufficient. Furthermore, we believe that a visual-like representation with spatial operations and reasoning techniques will be necessary for an image schema representation.

4 Computational models of image schemas and reasoning

The theory of image schema has been proposed for almost two decades, even some steps have been taken toward the computational formalization of image schemas, but image schemas are still largely discussed in qualitative, abstract terms.

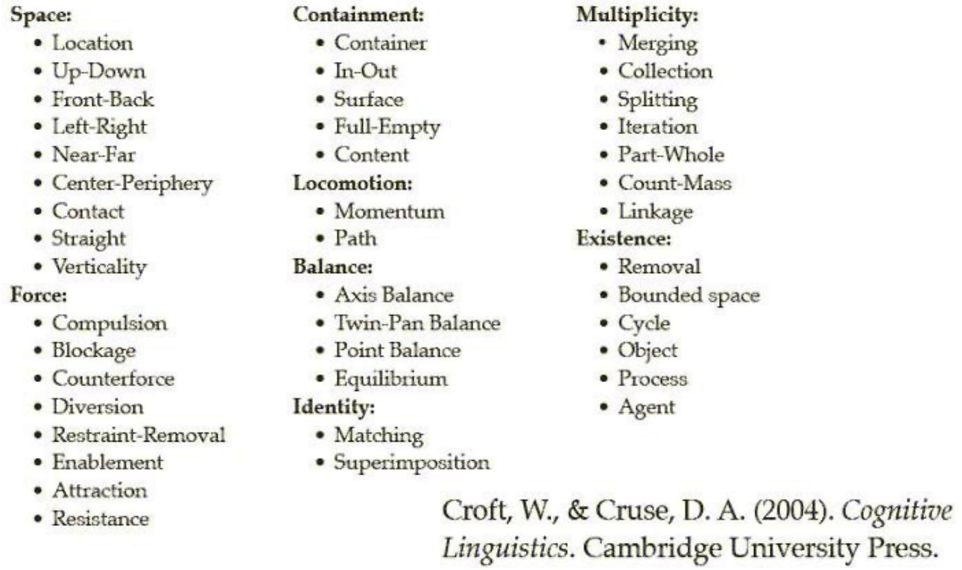
One explicit attempt to model image schemas can be found from Regier's work [Regier 1996]. Regier focused on spatial relations concepts, like in, out, from, to, on, off, front, back, above, below. He used a connectionist network and the following input: (1) project on a screen a set of "movies" consisting of a sequence of pictures of a geometrical figure either stationary or moving relative to others; (2) pair each movie with a spatial relation term in some natural language that correctly characterizes the spatial relation depicted on the input screen. Given such pairs of movies and terms, Regier has the network learn spatial relations concepts without negative examples, and test the result by having the network name new examples of spatial relations depicted in new movies.

Regier has developed what he calls "structured" or "constrained" connectionist neural models for a number of image schemas. "Constrained" neuro-computational connectionism builds into its neural models a small number of structures that have been identified in research on human visual and spatial processing. These include center-surround cell arrays, spreading activation, orientation-sensitive cells, and neural gating. Regier has shown how these constrained connectionist models of image schemas can learn spatial relations terms.

"If spatial relations are embodied rather than disembodied, the reason as a whole is also fundamentally embodied. Regiers work is a crucial step in showing that rationality itself arises from our bodily nature," writes George Lakoff.

Another computational formalization is called "Image Schema Language", a language in which image schemas can be modeled computationally, build by St.Amant et al. [2006]. The target of designing ISL is to let image schemas have a relational structure with compositional semantics that admits operations of interpretation and permits cross domain transfer of image schema structure. In another words, ISL represents image schemas in a general-purpose syntactic forms with the property that syntactic operations on them are equivalent to semantic operations in an indefinitely large number of domains.

The image schemas used in the paper is provided by Croft and Cruse (see figure 5). For each image schema, ISL represents it as an object with internal slots and set of operations. Slots are used to represent the rela-


Space:
- Location
- Up-Down
- Front-Back
- Left-Right
- Near-Far
- Center-Periphery
- Contact
- Straight
- Verticality

Force:
- Compulsion
- Blockage
- Counterforce
- Diversion
- Restraint-Removal
- Enablement
- Attraction
- Resistance

Containment:
- Container
- In-Out
- Surface
- Full-Empty
- Content

Locomotion:
- Momentum
- Path

Balance:
- Axis Balance
- Twin-Pan Balance
- Point Balance
- Equilibrium

Identity:
- Matching
- Superimposition

Multiplicity:
- Merging
- Collection
- Splitting
- Iteration
- Part-Whole
- Count-Mass
- Linkage

Existence:
- Removal
- Bounded space
- Cycle
- Object
- Process
- Agent

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.

Figure 5: Image Schemas by Croft and Cruse

tions with other objects. Operations determine the capabilities of an image schema object. For example, a container object may take one object as its content, one object as its boundary. It also has operations like “put into” and “get out”.

ISL has been implemented first in chess domain. Figure 6 demonstrates the relations of several image schemas in a chess board. On the top is the Path schema, which contains links and locations. Each location schema is represented by a container with a capacity of 1. When a piece uses a path schema to move, it needs to test whether the path is traversable, which is define by the status of every location in the path schema. Therefore, after a piece moves to a location, the container reaches capacity and yet another image schema, empty/full, is automatically created, indicating that the location is full, the path contains this location will not be traversable.

Later on, the work of ISL move to a more complicated domain, a “playpen” environment with ballistic physics, modeled in the Breve 3-D simulation engine [Klein 2003], as shown in figure 7. In this domain, an agent (blue rectangle) is learning to catch the target (red ball). The target can be either a ball, which means no initiative actions, or a cat, which can be threatened

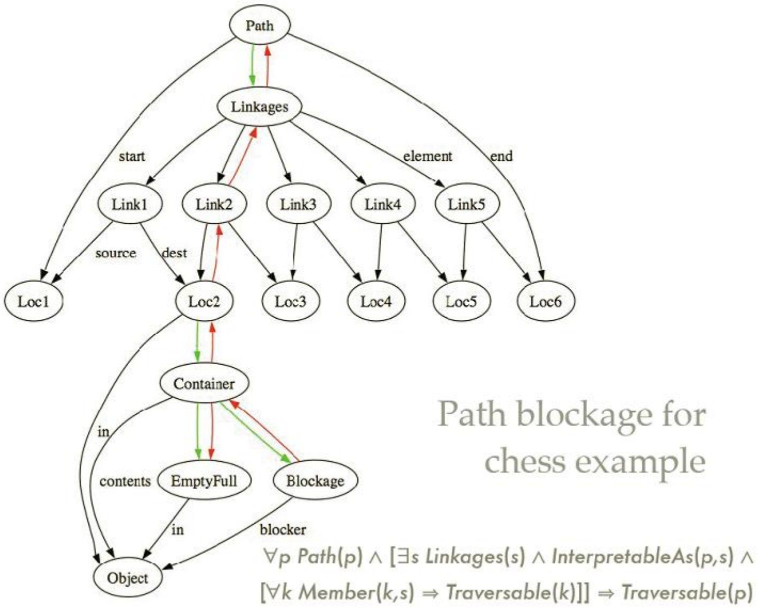


Figure 6: ISL reasoning in chess board

and run away from the agent. The cat get alert only if the agent move too fast or too close. Given this behavior, one strategy for the agent to catch the cat is to move to the cat slowly (sneak), and then move very rapidly to the cat (pounce) once the cat is close.

Using ISL, a state machine by which the agent understanding the world can be build (see figure 8). For each state, it contains the set of image schemas describing the current situation. The transition represents the action the agent takes. Therefore, if the target only can be a ball, only two states (S1, S5) are enough. But when the agent is chasing a cat, the ISL description of the current situation will not be consisted. In such case, a state split algorithm will be called, and new state will be generated based on the entropy. After enough examples have been trained, a state machine which combines the cats behavior will be generated.

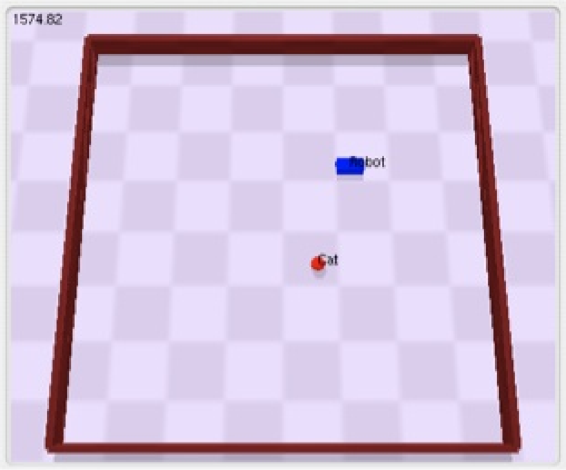


Figure 7: Breve simulation

5 Combination of depictive and propositional representation and reasoning

5.1 Trace-map: A prototype representation

As discussed above, a pure propositional representation does not appear to support the kind of interpretation and reasoning associated with image schemas. In this section, we outline an approach that shows how a representation can be built from the sensor input and motor actions of a simulated embodied agent. This representation includes both depictive and propositional components, and it can be used to support some simple processing that can be interpreted in image schematic terms. Our discussion is incomplete, but we believe that the direction is promising.

Figure 9 is a simple simulation of a robot arm. Both pink and light green part can be rotated around the joint. Assume it has the following configuration (10,7) — the length of each part. We have four controllers associated with this arm. They are (0, 1): rotate the lower arm CCW (counterclockwise); (1, 0): rotate the upper arm CCW; (1, 2): rotate both arms CCW, both lower one move twice fast as the upper one; (1,-2): rotate upper CCW, lower CW (clockwise), and the speed ratio is 1:2. We can record the position of the robot arms endpoint as shown in the figure 10.

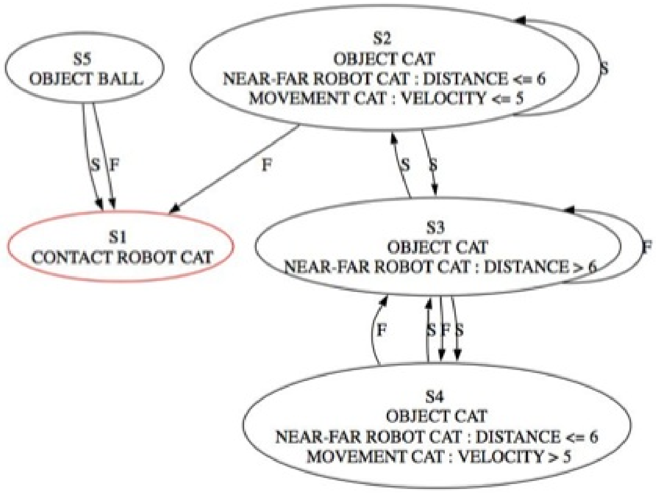


Figure 8: State machine of the simulation

Although the actual sizes and positions are different, people can identify some similarity among these maps. This gives a hint about how to store the information in a unified way. A unified way to store those trace-maps should be,

- Store a pixel by pixel map of a given shape in the same uniform size[3].
- Present the given image as large as possible in trace-map.
- Minimize the over all y-value of the trace as low as possible.
- Record the parameters used in the above.

The motive of these rules is to keep consistency of trace-maps under translation, scaling, and rotation. For each trace, the parameters will look like figure 11.

There are several reasons we want this representation. First, similar image can be represented under the same trace-map. We can implement a function to compute the similarity of different maps[4]. Maps like (0, 1) (1,

---

[3] I use a 30*30 matrix in my prototype.

[4] Even a simple function like comparing maps pixel by pixel can give us a pretty good illustration.

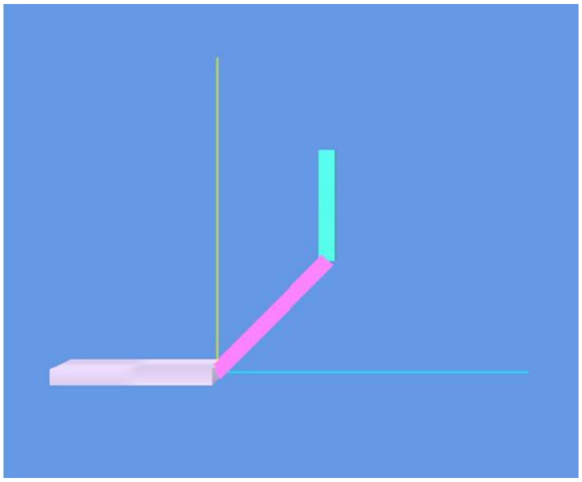


Figure 9: A simulated robot arm

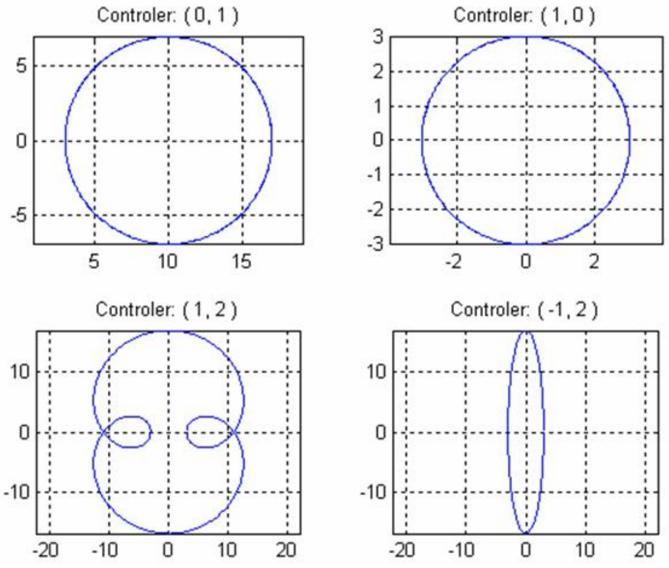
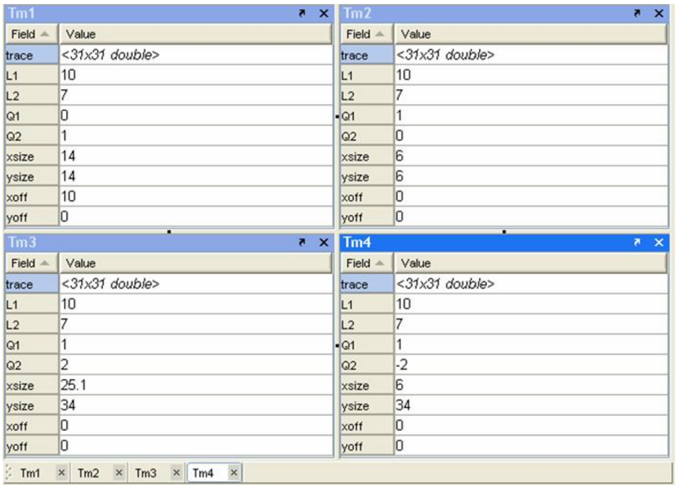


Figure 10: The controller's traces

0) (-1, 2) are highly related, but not for (1, 2). However, similar configuration of the robot arm will lead to the similar maps. Especially for the complicated map like (1, 2), its very interesting to see it can be used for configurations like (10, 5), (10, 6) and (10, 8) (see figure 12). Their topological properties and some of their quantitative properties remain the same.



Tm1

Field	Value
trace	*<31x31 double>*
L1	10
L2	7
Q1	0
Q2	1
xsize	14
ysize	14
xoff	10
yoff	0

Tm2

Field	Value
trace	*<31x31 double>*
L1	10
L2	7
Q1	1
Q2	0
xsize	6
ysize	6
xoff	0
yoff	0

Tm3

Field	Value
trace	*<31x31 double>*
L1	10
L2	7
Q1	1
Q2	2
xsize	25.1
ysize	34
xoff	0
yoff	0

Tm4

Field	Value
trace	*<31x31 double>*
L1	10
L2	7
Q1	1
Q2	-2
xsize	6
ysize	34
xoff	0
yoff	0

Figure 11: The parameters of trace-maps

Second, the relations between the map and the "real-world" have been carried out by the trace-map's parameters. Like controller (0, 1) and (1, 0), although they have the same shape, but the position and the size is decided by the endpoint of robot arms. This relationship can be represented just by the parameters of the trace-map, it's much easier to discover than before. That is, we have a concise representation in which similarities, within certain geometrical and topological constraints, can be determined in a computationally efficient way.

Within a trace-map, not only the topological information is stored, but also the orientation lead to the next point by executing the current controller can be record. Figure 13 demonstrate this information — the actual height represents the angle of the movement at the point. This information also can be used in different ways. For example, it can help us to distinguish the clockwise orientation and counterclockwise orientation. And it's also useful when a planning algorithm is applied, like what we discuss in the next paragraph.

The last and most important usage of trace-map is it can be treated like a flexible experience and applied dynamically. In this robot arm experiment, a result of using one controller could be the same result if you apply the map back to the situation. For example, suppose we have the relations build with

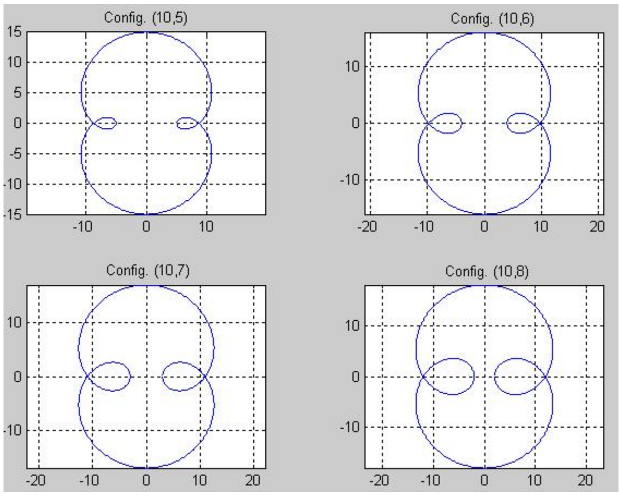


Figure 12: The variation of configurations

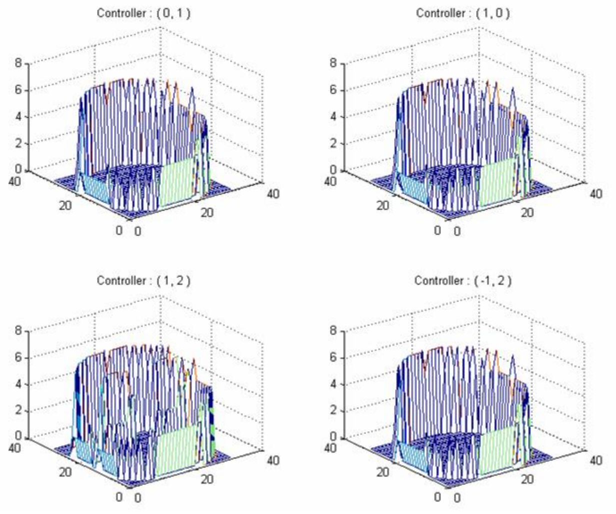


Figure 13: Direction information of trace-maps

those maps. What if we need to reach the red point in figure 14? We can apply the trace-map like (0, 1) and find the closest point on the circle. This will give us one choice. By comparing all the trace-maps we have, we can choose our controller easily by a greedy algorithm. Of course we could have better solution, but the point here is to give us a better way to use planning or search algorithms.

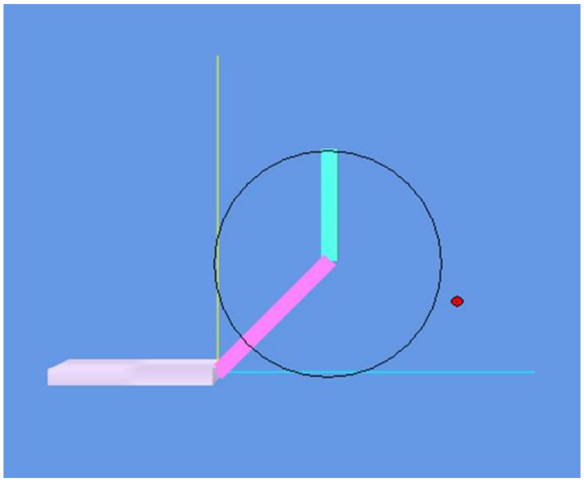


Figure 14: Apply trace-maps in planning

5.2 Developing the mixed representation

The idea behind the trace-map is to combine the depictive and propositional representations. A map obviously can be used as a depictive representation; it records information including shapes, spaces, topological relations and moving directions. In order to meet the requirements of image schemas and mental imagery, geometric operations and spatial reasoning also need to be the part of the map functions. For example, in Hegarty's gear rotation case, if we have separate maps representing the gears, we can apply rotation directly to have a mental imagery.

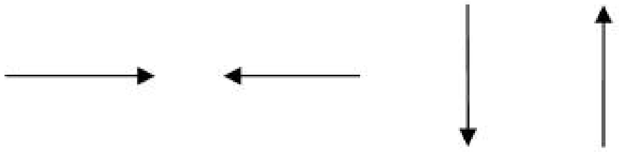


Figure 15: PATH example 1

As Kosslyn pointed out, for a depiction to have meaning, it must be interpreted in a specific way. A map also needs to be associated with propositional representations. The parameters we used in the trace-map representation record its geometry information (size, position, orientation), to specify where and how to use a trace-map. Some meanings can be represented in this way. For example, the most natural interpretation of trace-map is a PATH schema. Consider the representation of the dynamic behavior of a trace map for a controller such as (0,1). Each of the transitions represented in the discrete map can be considered a path or a path component. Furthermore, image schemas like LEFT-RIGHT, UP-DOWN can be represented in fingure 15[5].

Similarly, the FAR-NEAR relation can be known by comparing the scale parameter. Another complicated meaning is by comparing the shape (a simple straight arrow and the arrow like the following), the map could show the difference between the static PATH and the animate PATH (figure 16).

Figure 16: PATH example 2

However, the combination of map and parameters alone is not enough. There are lots of static relations in image schemas that need to be specified. For example, a container schema partitions the space into three parts, outside, inside and the boundary. In this case, a static map with marks on different parts is needed (figure 17). Here, we use the map in another way, as a static-map. In this kind of map, we don't need to specify the trace of a movement; instead, we only need to characterize the shape and different marks associated with this shape.

The static-map also needs to be interpreted by propositional representations. The function of this representation is to capture relationships. For example, a static-map above need a structure to say area 1 and area 0 has IN-OUT relations with a circle boundary in between. Structures used by image schema language express the same idea, but without a pictorial description. Including the static-map as a part of an image schema structure

---

[5] Actually, after we unify the maps, its the orientation parameter that distinguishes them

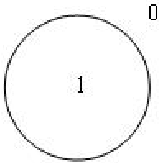


Figure 17: A static-map of container

will give it more power, like qualitative spatial reasoning [Cohn and Hazarika 2001].

The difference between trace-map and static-map is that one expresses dynamic schemas and another one expresses static schemas. For a trace map, the trace of a movement is important and should be record point by point. For a static map, the distinction of different parts in the map is important and could be record by different marks.

Another important concept we use for image schemas is hierarchical structures. An image schema can be composed of more basic image schemas, so that a more complicated meaning can be represented. For example, by combining a static-map and a trace-map, we can express meanings like "move from inside to outside", "move within the boundary", "blocked by the boundary", and "move from left to right", etc (figure 18).

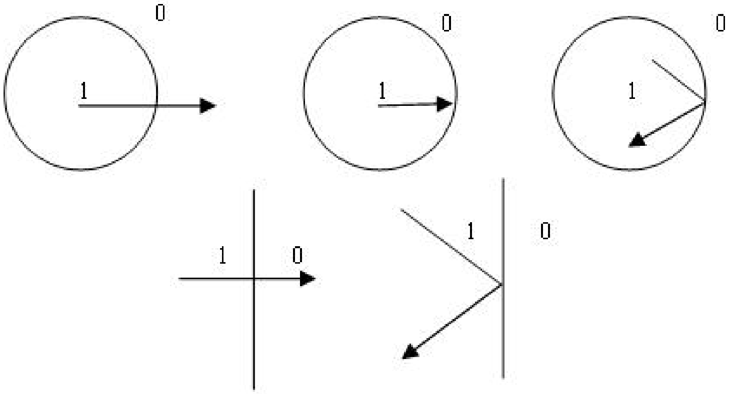


Figure 18: Examples of map combination

Multiplicity schemas also can be expressed in hierarchical structures. The PART-WHOLE schema is a good example. In last section, the controller (0,1) trace we used is a circle. But if we focus on a small part, and use a single trace-map to represent it, it will look like figure 19. In this case, we can consider it a LEFT-RIGHT movement, or an animation movement, etc. These are relations we can contain within a PART-WHOLE schema.

Figure 19: Part of a circle trace-map

Although a lot of technical details remain unknown, associating the maps with propositional representation seems a plausible way to extend the capability of traditional representations. And many people have done research work in the similar ways. For example, Cohen [1998] has described maps for verbs, which by considering the trajectories in map it can distinguish the meanings of two objects' movement, like “approach”, “escape”, “hit” etc. Other research done by Yu-Han Chang refines basic maps from sequences of continuous observations. He designed an agent using maps to predict a target‘s behavior. When an unpredicted situation happened, he tried to split the map into several sub-maps by an entropy measure to find the causal relationships.

To summarize, our further work will be involve implementing the idea of this associated map into a tool-using agent. We hope by this way, we can combine the power of depictive and propositional representations, and we can build an agent that has some properties implied by the theories of image schemas and mental imagery.

References

Sven Bertel, Thomas Barkowsky, Dominik Engel, and Christian Freksa. Computational modeling of reasoning with mental images: Basic requirements. In *Proceeding of the 7th International Conference on Cognitive Modeling*, 2006.

Paul R. Cohen. Maps for verbs. In *Proceedings of the Information and Technology Systems Conference, Fifteenth IFIP World Computer Congress*, pages 21–33, 1998.

A. G. Cohn and S. M. Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta informaticae*, 46(1):32, 2001.

Mary Hegarty. Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, 8(6):280, 2004.

Mark Johnson. *The body in the mind : the bodily basis of meaning, imagination, and reason.* University of Chicago Press, Chicago, 1987. ISBN 0226403173.

Mark Johnson and Tim Rohrer. We are live creatures: Embodiment, american pragmatism, and the cognitive organism. *Body, Language, and Mind*, 1, 2006.

Jon Klein. breve: a 3d environment for the simulation of decentralized systems and artificial life. In *ICAL 2003: Proceedings of the eighth international conference on Artificial life*, pages 329–334, Cambridge, MA, USA, 2003. MIT Press. ISBN 0-262-69281-3.

Stephen Michael Kosslyn. *Image and Brain : The Resolution of the Imagery Debate.* The MIT Press, Cambridge, Massachusetts, 1996. ISBN 0-262-61124-4.

George Lakoff and Mark Johnson. *Metaphors we live by.* University of Chicago Press, Chicago, 1980. ISBN 0226468011.

George Lakoff and Rafael E. Núñez. *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics Into Being.* Basic Books, New York, NY, 1st ed edition, 2000.

Todd Oakley. Image schema. In Dirk Geeraerts and Hubert Cuyckens, editors, *The Handbook of Cognitive Linguistics.* Oxford University Press, Oxford and New York, In press.

Zenon Pylyshyn. Return of the mental image: are there really pictures in the brain? *Trends in cognitive sciences*, 7(3):113, 2003.

Terry Regier. *The Human Semantic Potential: Spatial Language and Constrained Connectionism.* The MIT Press, Cambridge, Massachusetts, 1996. ISBN 0-262-18173-8.

Tim Rohrer. Image schemata in the brain. In Beate Hampe and Joe Grady, editors, *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, pages 165–196, Berlin: Mouton de Gruyter, 2005.

Robert St.Amant, Clayton T. Morrison, Yu-Han Chang, Wei Mu, Paul R. Cohen, and Carole Beal. An image schema language. *The 7th International Conference on Cognitive Modeling*, 2006.