Incorporating Visual Imagery into a Cognitive Architecture

Theory, Design, and Implementation

Scott Lathrop

John Laird

26th SOAR WORKSHOP

OUTLINE

- BACKGROUND & MOTIVATION
- ARCHITECTURE
- TABLE SETTING DOMAIN
- GEOMETRY PROBLEM DOMAIN
- NUGGETS & COAL

WHAT IS VISUAL IMAGERY?

- **What is larger, a softball or a baseball?**

WHAT IS VISUAL IMAGERY?

- **What hand does the Statue of Liberty hold the torch?**

WHAT IS VISUAL IMAGERY?

- **What city is closer to Ann Arbor: South Bend, Indiana or Columbus, Ohio?**

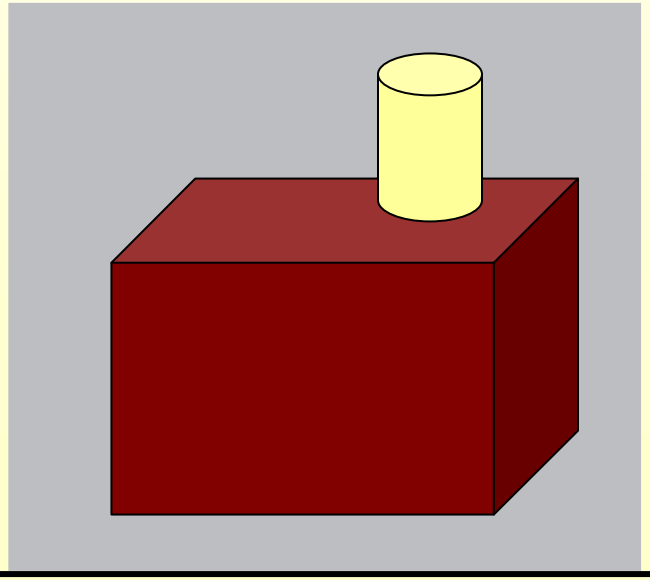
WHAT IS VISUAL IMAGERY?

Symbolic (Descriptive)	Visual Image (Depictive)
Explicit, abstract objects	Implicit, concrete, objects
Explicit relations	Implicit relations
Location, size, shape, features, orientation optional	Location, size, shape, features orientation inherent
Computationally efficient for maintaining semantic interpretations	Computationally efficient for maintaining visual and spatial relationships

Symbolic

Object (can)
Yellow(can)
Object(box)
Red(box)
On(can, box)

Visual Image



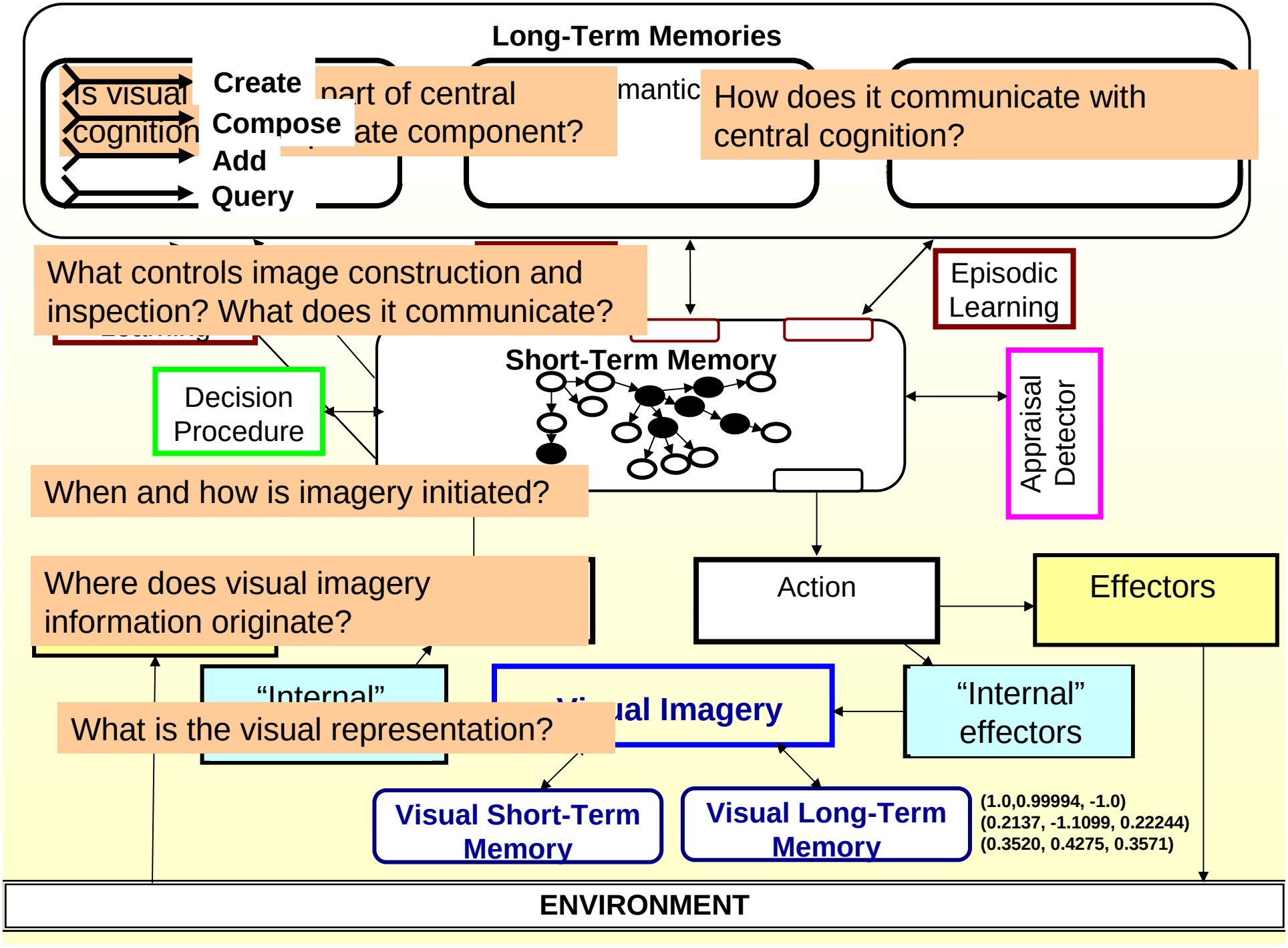
*Adapted From Kosslyn Image and Mind

WHY STUDY VISUAL IMAGERY?

- "Best of both worlds" multi-modal approach
- General AI Agent/Cognitive Architectures
  - Symbolic representations & computations
  - Fewer efforts to integrate sensory modalities
- Depictive Representations
  - Visual and spatial format
  - Computationally more efficient for visual-spatial tasks.
  - Requires less domain knowledge for visual-spatial tasks
- Applicable to visual-spatial domains

RESEARCH GOALS

- To incorporate visual imagery within the context of a cognitive architecture constrained by psychological and biological evidence
- To improve the spatial reasoning capability of Soar
- To understand visual imagery's functional capabilities and limitations
  - What are the environment and task conditions where visual imagery provides additional capabilities to a general cognitive architecture?
  - Under what environment and task conditions is it computationally more efficient than a symbolic representation?
  - When does it require less task knowledge?


Long-Term Memories
Create
Compose
Add
Query
Is visual cognition part of central ate component?
How does it communicate with central cognition?
What controls image construction and inspection? What does it communicate?
Episodic Learning
Short-Term Memory
Decision Procedure
Appraisal Detector
When and how is imagery initiated?
Where does visual imagery information originate?
Action
Effectors
"Internal"
ual Imagery
"Internal" effectors
What is the visual representation?
Visual Short-Term Memory
Visual Long-Term Memory
(1.0,0.99994, -1.0)
(0.2137, -1.1099, 0.22244)
(0.3520, 0.4275, 0.3571)
ENVIRONMENT

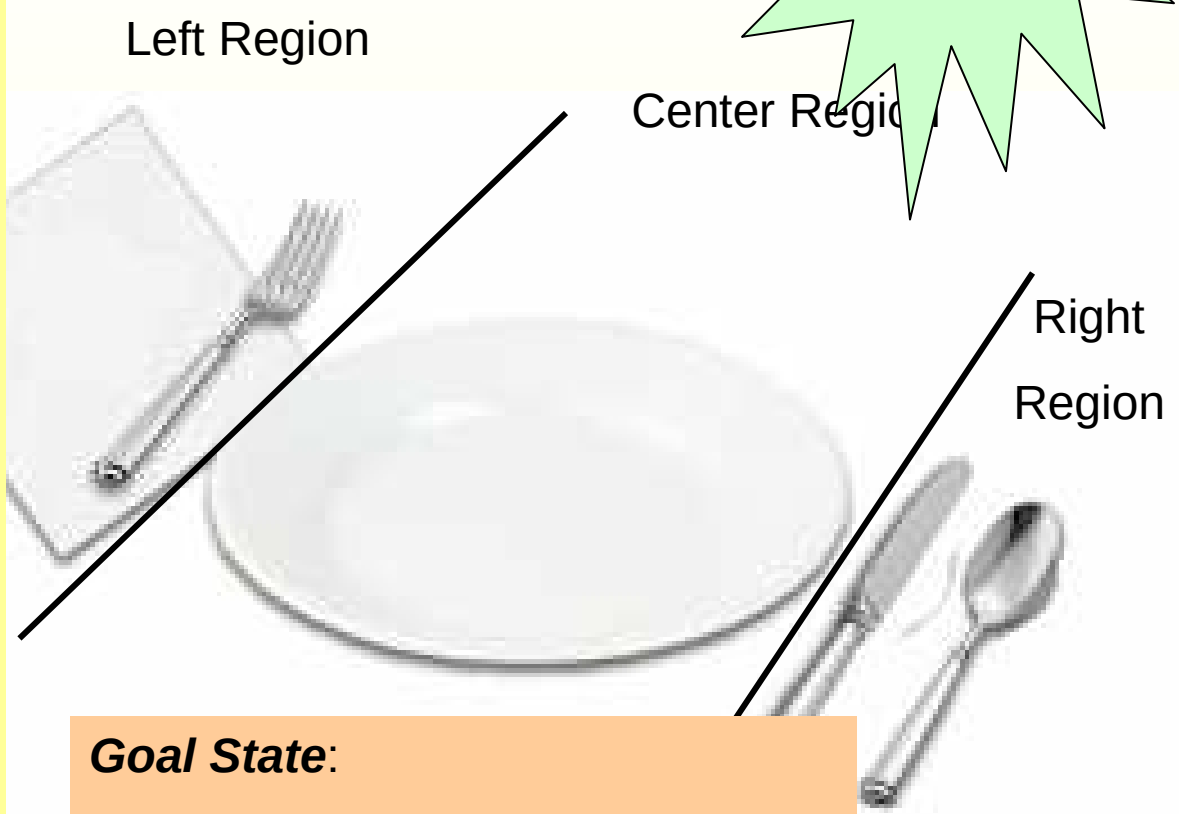
EXPERIMENT # 1
TABLE SETTING

What is the center object?

Initial Knowledge

- Fork, Napkin, Plate, Knife, Spoon
- Place Setting made up of above objects
- Local Relationship
  - On-top(Fork, Napkin)
  - Left-Of (Fork, Plate)
  - Right-Of (Knife, Plate)
  - Right-of (Spoon, Knife)

- *No knowledge of global relationships*



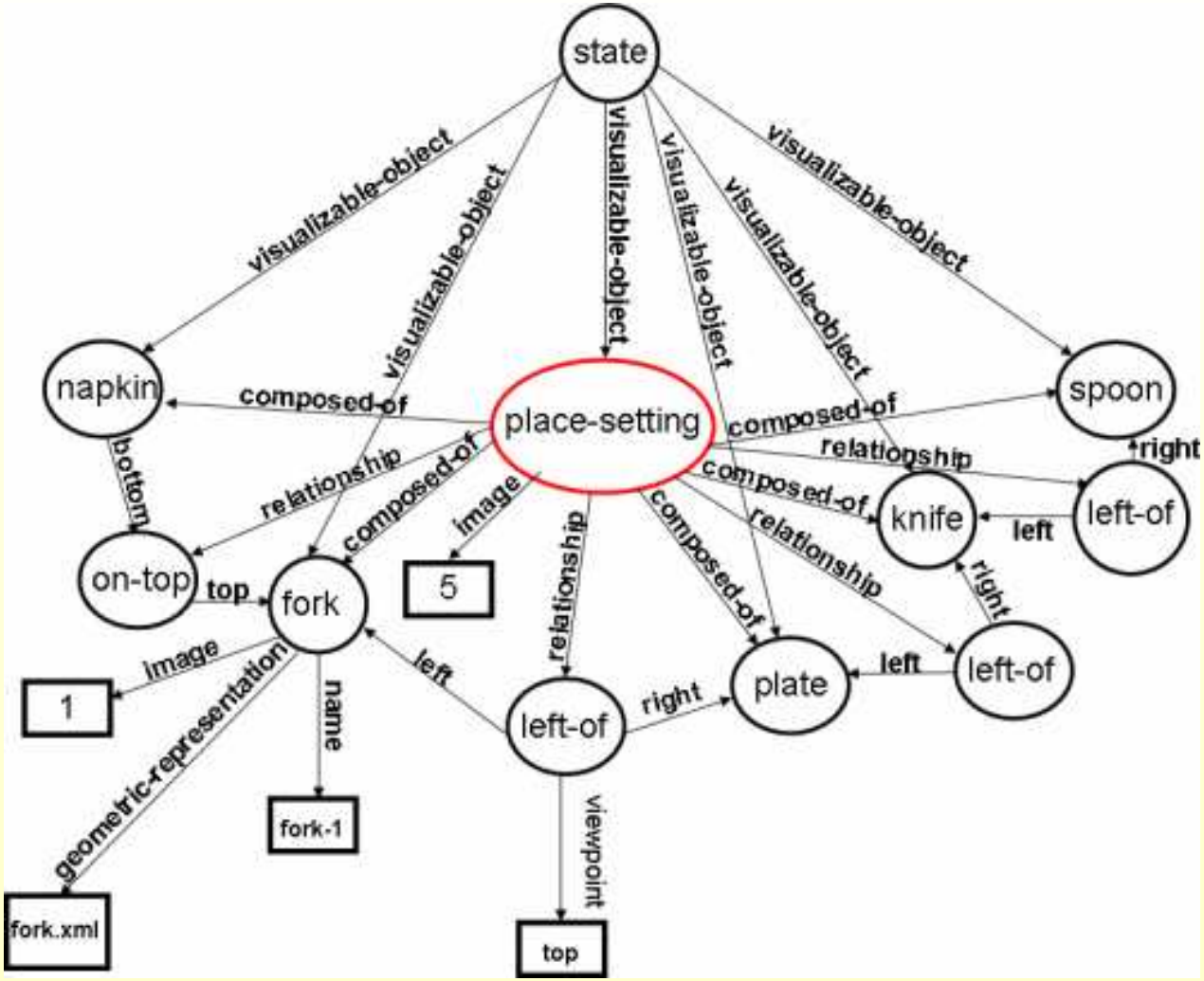
Goal State:

Left Region: Fork above Napkin

Center Region: Plate

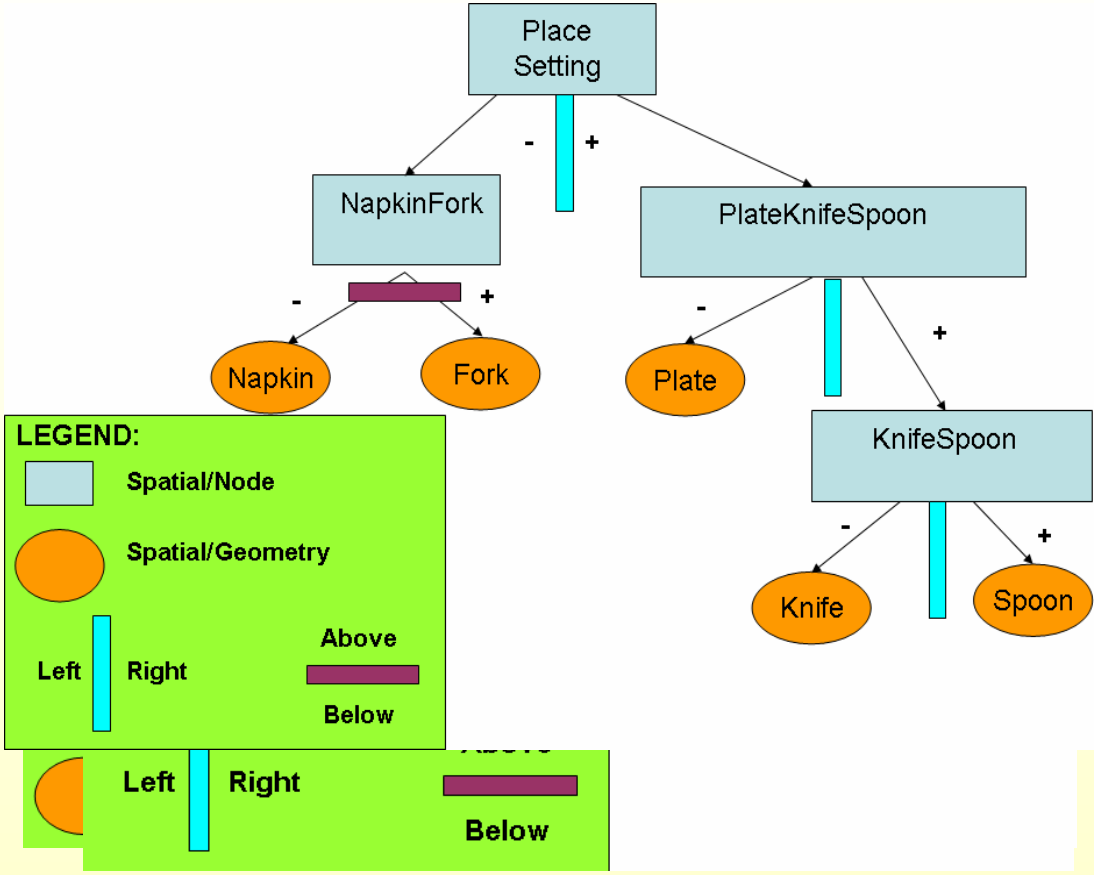
Right Region: Knife left of spoon

SYMBOLIC REPRESENTATION



VISUAL REPRESENTATION

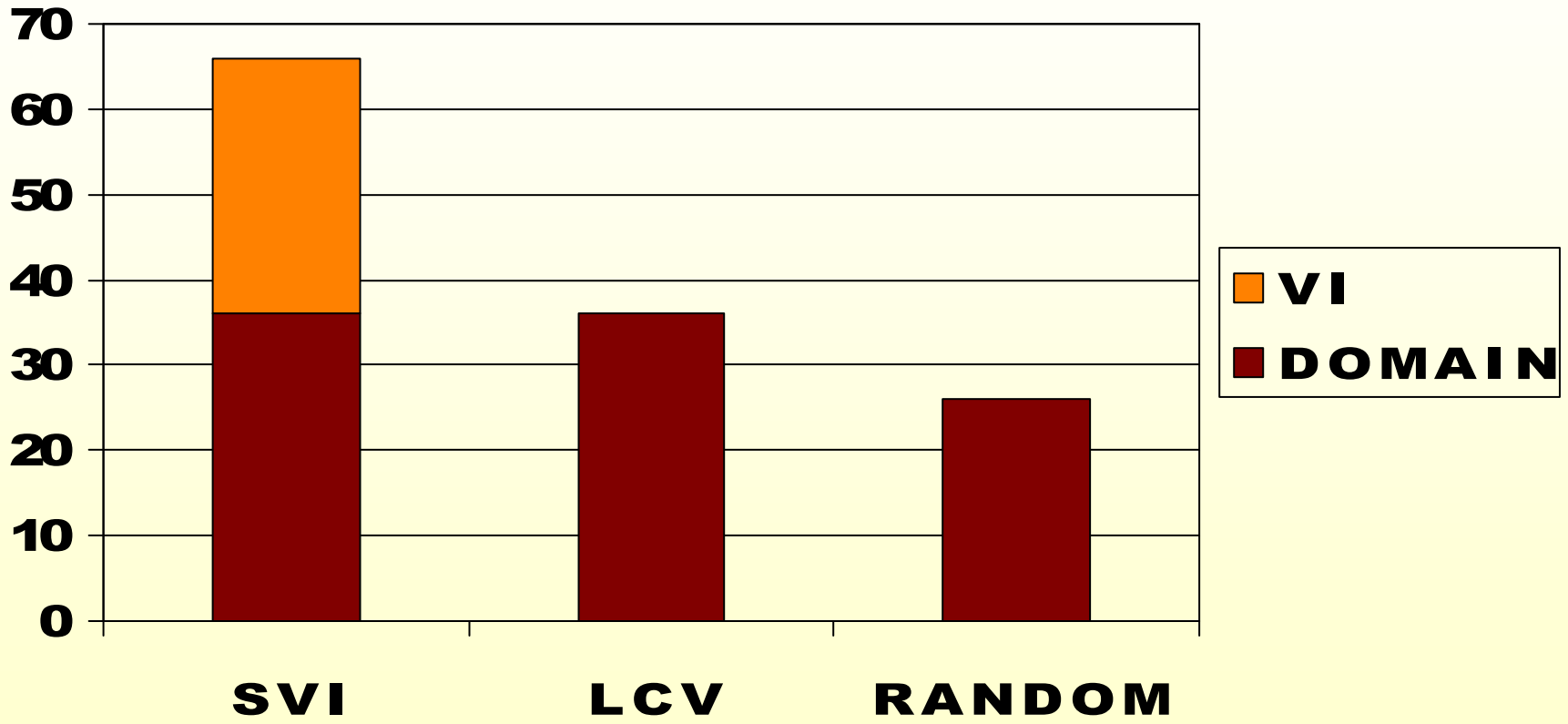
- **Scene Graph / BSP Tree**
  - DAG
  - Represents a region of space
- **Desired Properties**
  - Hierarchical
  - Combines Geometric and Spatial data
  - Logical and Spatial Groupings (ideally)
  - Renderable to 2D image



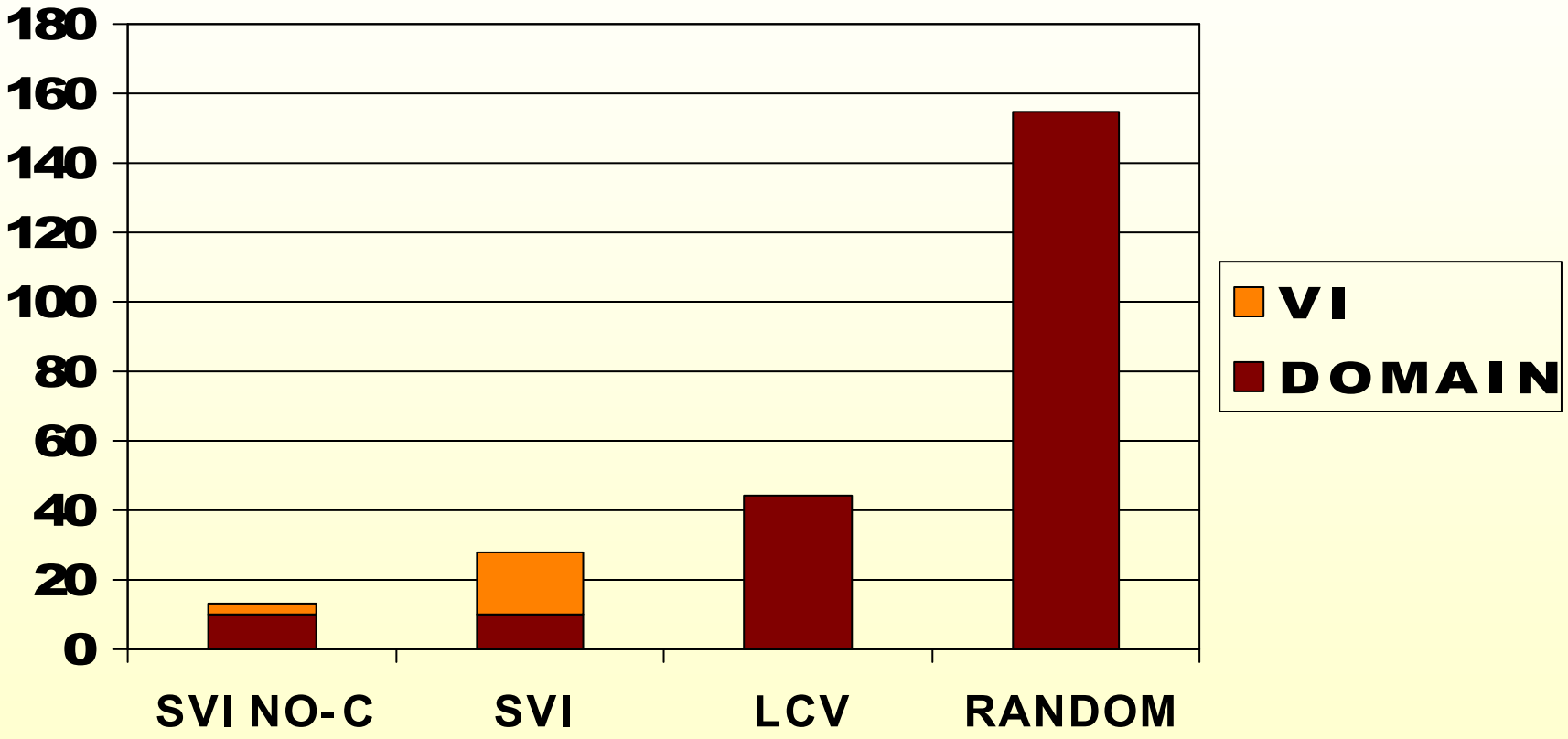
EVALUATION CRITERIA

- **Knowledge – Number of productions (Long term knowledge)**
- **Problem Solving – Number of decision cycles (count of reasoning steps in Soar)**
- **Computational Efficiency**
  - Time in Soar kernel (central cognition)
  - Time in Visual Imagery
  - Total Time
- **Functional Capability – What additional capability does visual imagery provide?**
- **Constraints – Must work with given behavioral, structural, functional, and computational constraints**

PRODUCTION COUNT

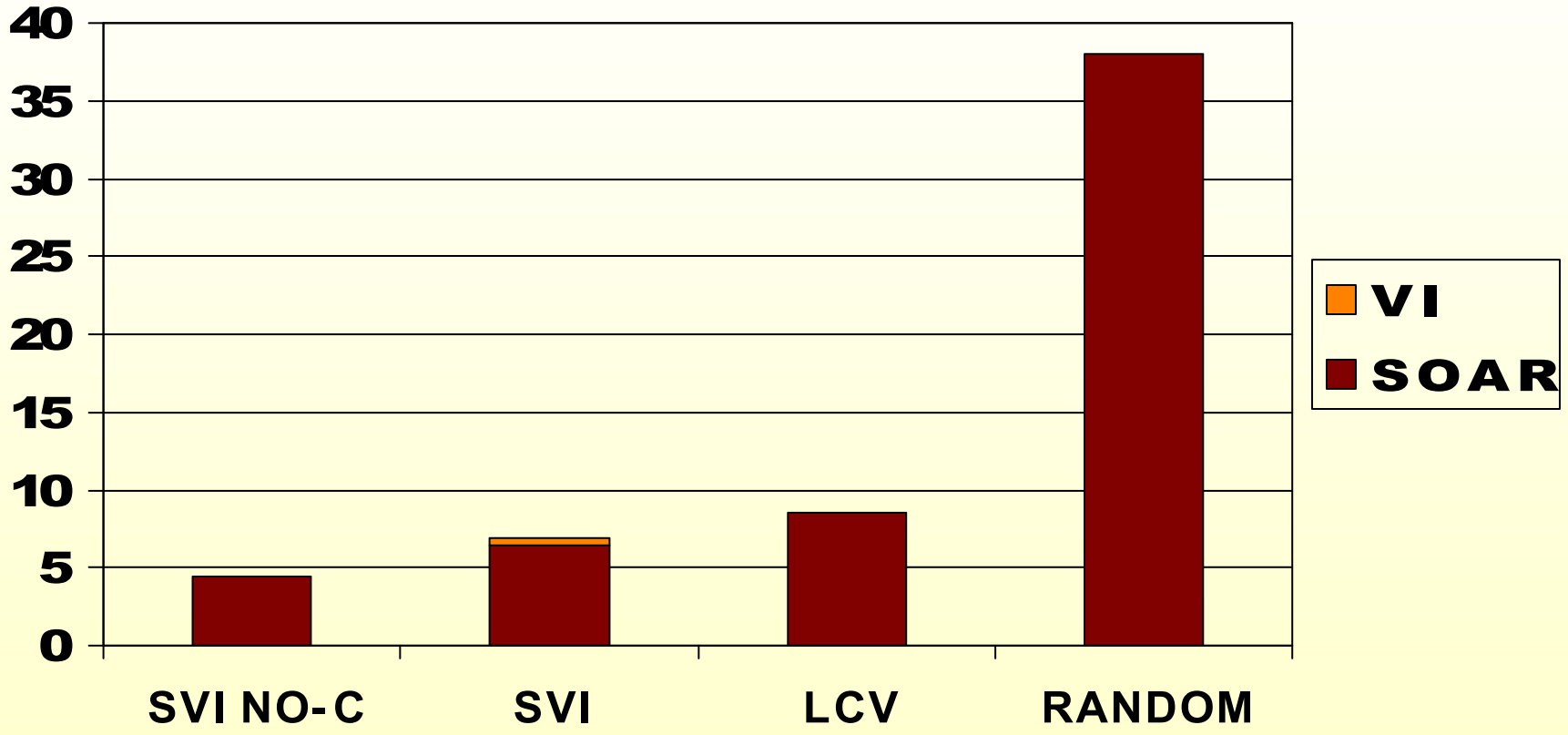


DECISION CYCLES



* Median over 30 trials

TIME (ms)



* Median over 30 trials

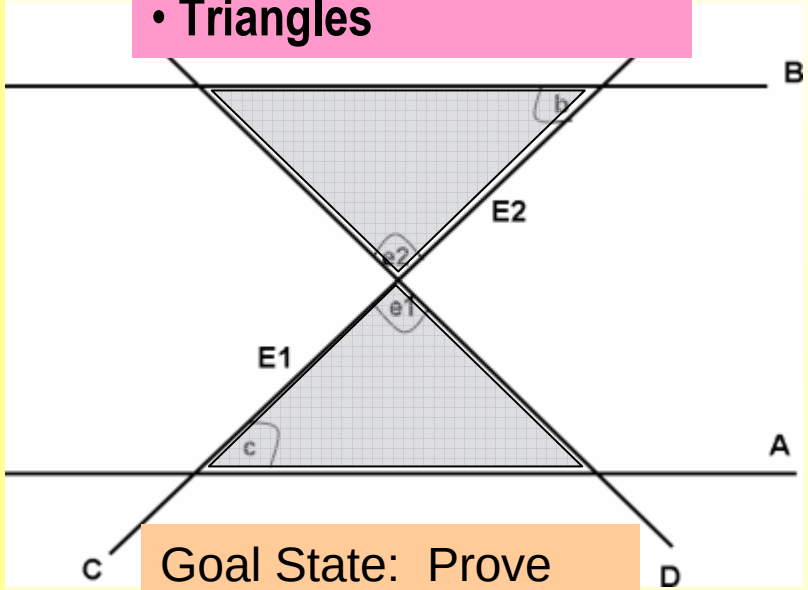
EXPERIMENT #2: GEOMETRY PROBLEM

Initial Knowledge

- 4 lines (A, B, C, D)
- A is parallel to B
- C intersects A
- D bisects the line segment formed by the intersection of C and A and C and B
- If something is a bisector, then it divides line segment into two congruent segments
- If two angles are alternate interior angles, then they are congruent
- If two angles are vertical angles, then they are congruent
- (ASA Rule). If two angles and the included side of one triangle are congruent to the corresponding two angles and included side of another triangle, then the triangles are congruent

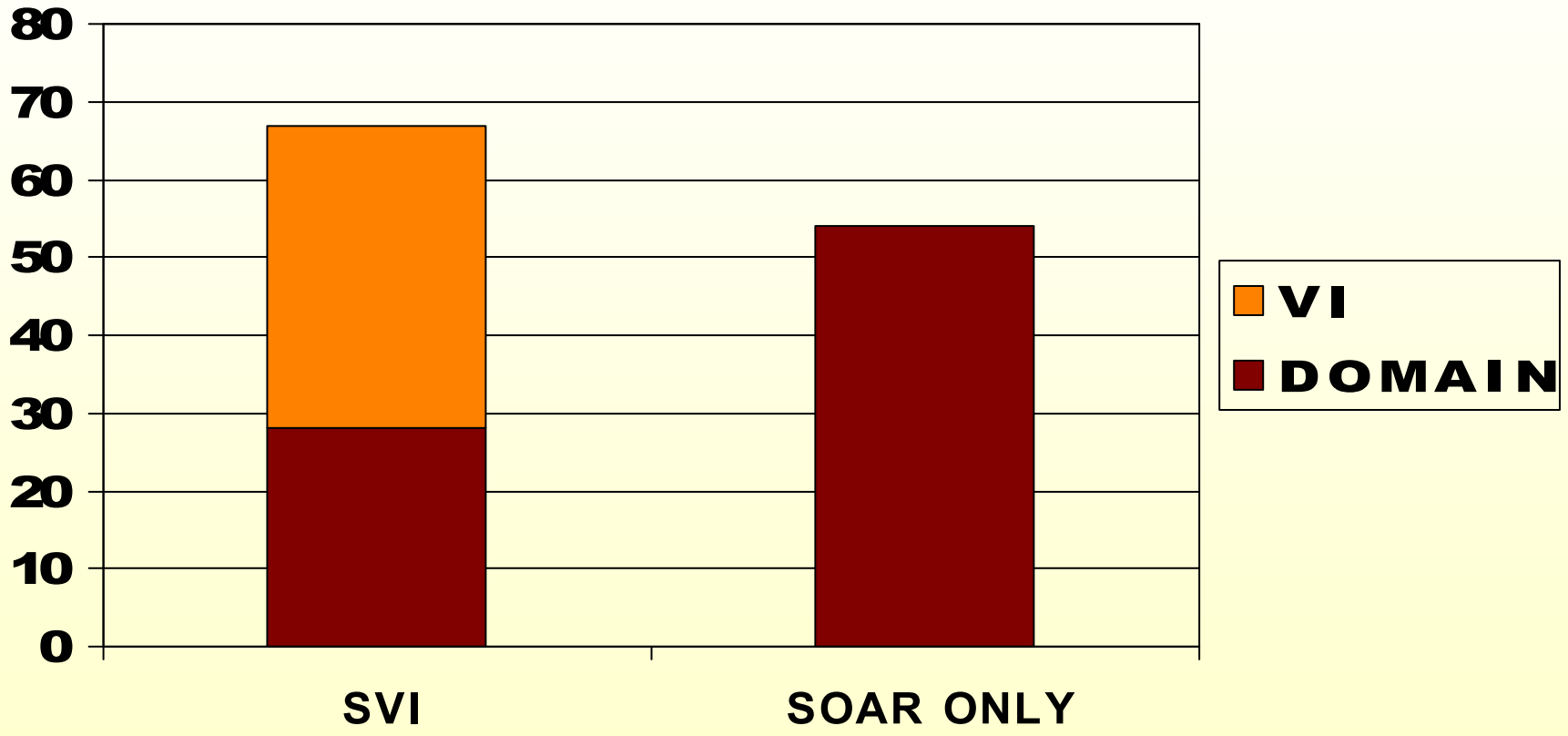
What's Missing

- Angles
- Points
- Line Segments
- Triangles

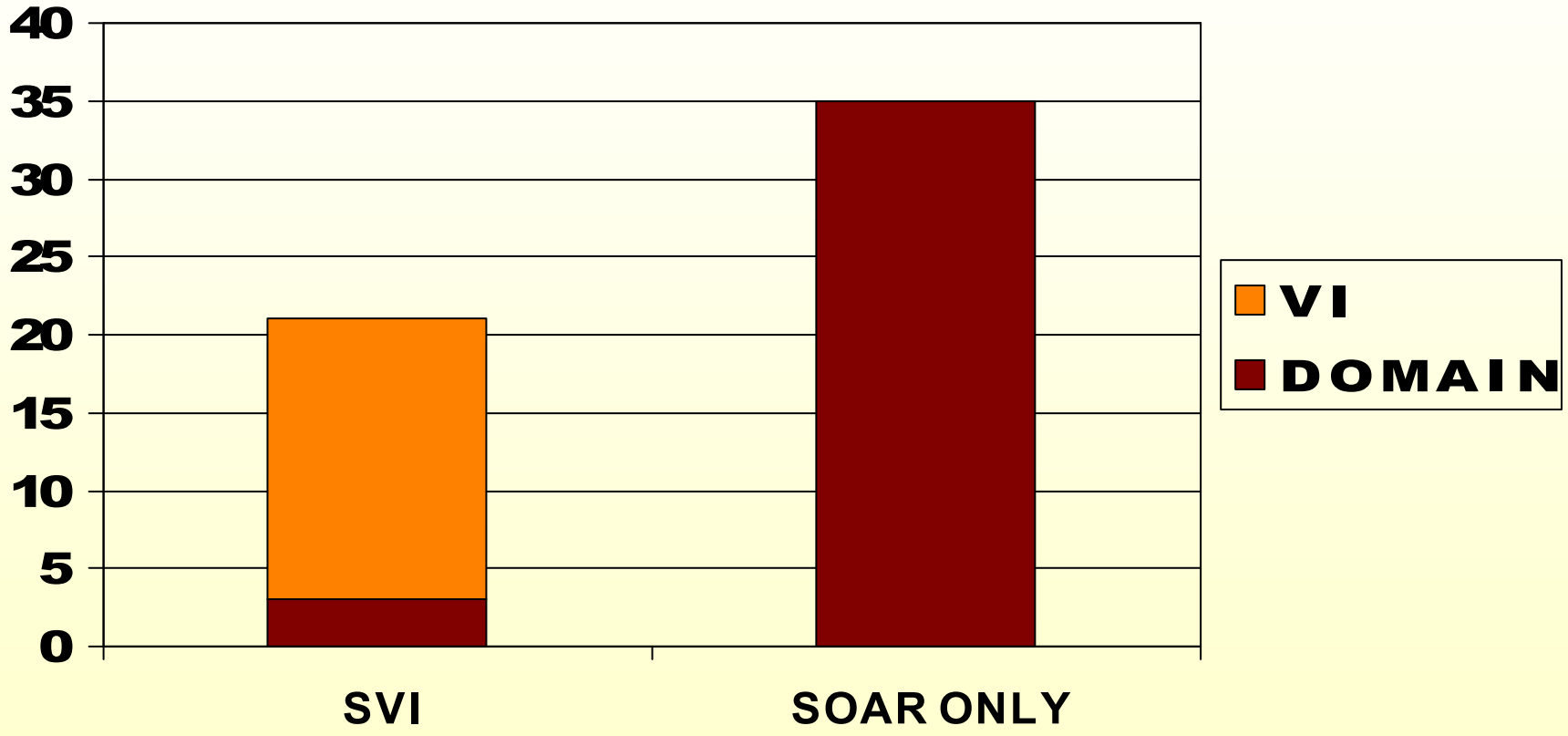


Goal State: Prove the two triangles congruent

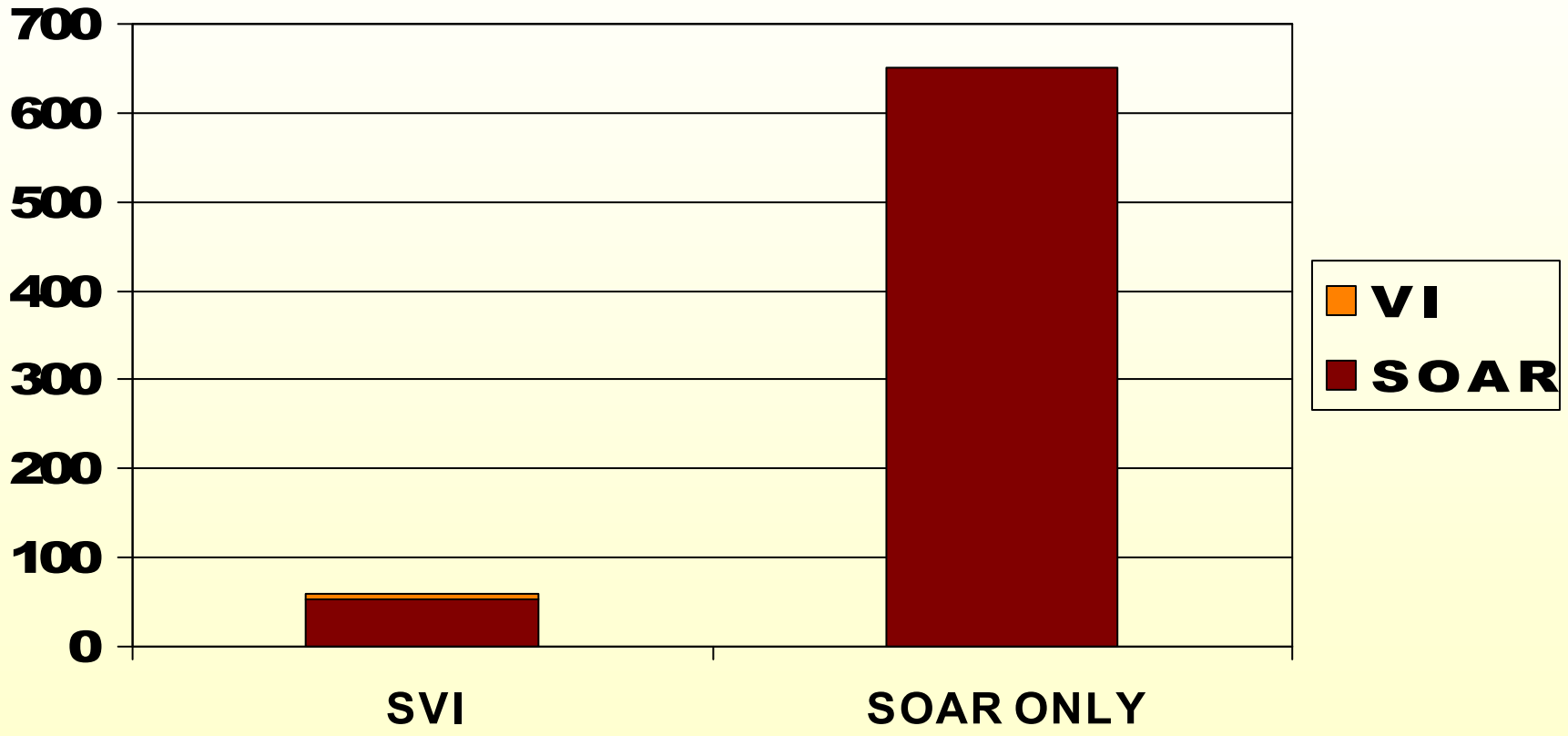
PRODUCTION COUNT



DECISION CYCLES



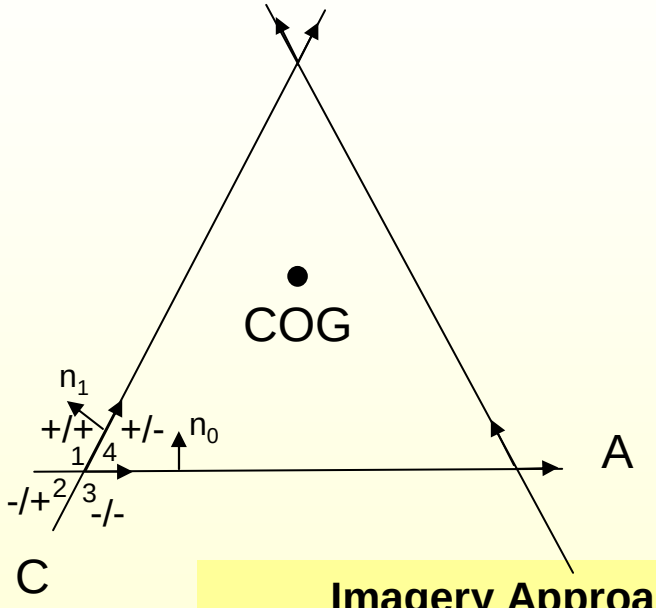
TIME (ms)



* Median over 30 trials

FUNCTIONAL CAPABILITY

- **"Sense of direction"**
  - **Required to find triangle's interior angles**
  - **Symbolic approach alone not capable without providing more knowledge**



Symbolic Approach

- One angle per region (1,2,3,4)
- Which angle belongs to the triangle?
- Which side of the line does each angle belong to?

Imagery Approach

- Find triangle center of gravity (COG)
- For each triangle line segment, determine side COG is on (+/-).
- Triangle angle at each vertex corresponds to region (+/+, -/+, -/-, +/-)

KEY POINTS

- Central Cognition
  - Domain knowledge (problem space, goals, states)
  - Controls visual image construction and inspection
- Visual Imagery
  - Constructs and searches what it is "told"
  - Provides perceptions based on what it "sees"
    - Visual properties
    - Spatial relationships
  - Is not solving problem (it has no notion of problem space, goals, states, etc.)
  - Performs efficient computations and provides new knowledge
  - Reacquires concrete knowledge "abstracted away" during initial perception

NUGGETS & COAL

- **NUGGETS**
  - Depictive representations show promise for achieving computational efficiency in visual-spatial domains
  - Additional functional capabilities for spatial reasoning

- **COAL**
  - Still looming questions as to what is the image's internal representation? Is there more than one format/data structure?
  - Lot of unknowns about the relationships between vision and visual imagery
  - Truth maintenance between central cognition, imagery, and vision

QUESTIONS

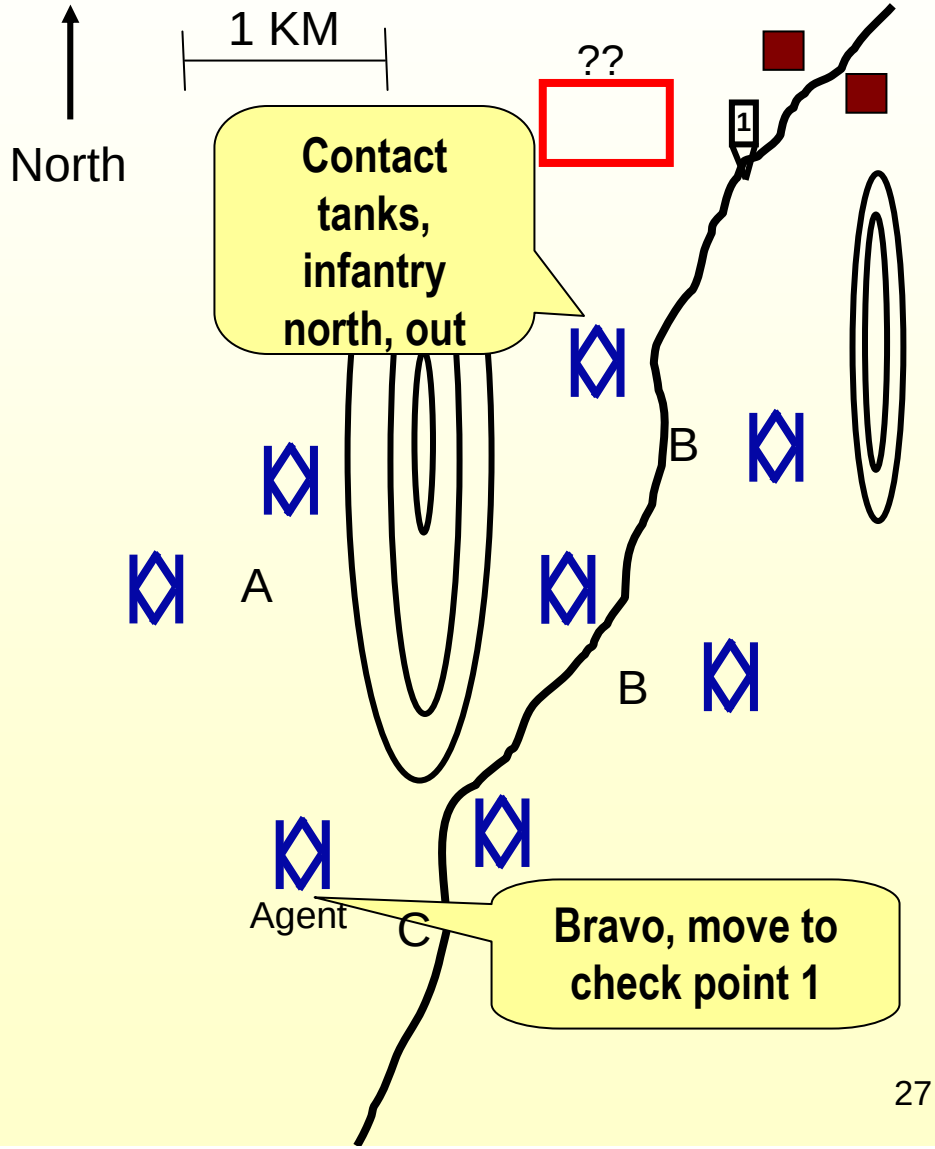
BACKUPS

BACKGROUND & MOTIVATION

- **Why study visual imagery?**
- **What is visual imagery?**
  - Structure and processes that construct and interpret a *visual representation*
  - Computational efficient in visual-spatial tasks
  - *Assists us in* reasoning, planning, learning, and motor skills
- **What are potential applications?**
  - Simulation/Games
  - Robotics
  - Decision Support Tools
  - Intelligent GIS

COGNITIVE CAPABILITY VISUAL SIMULATION EXAMPLE

- Agent is “scout platoon leader”
- Task: Reconnaissance
- Sub task: gain visual contact
- Goal: To establish enemy force’s breadth & depth



Legend:

Scout Vehicle

Enemy Force

??

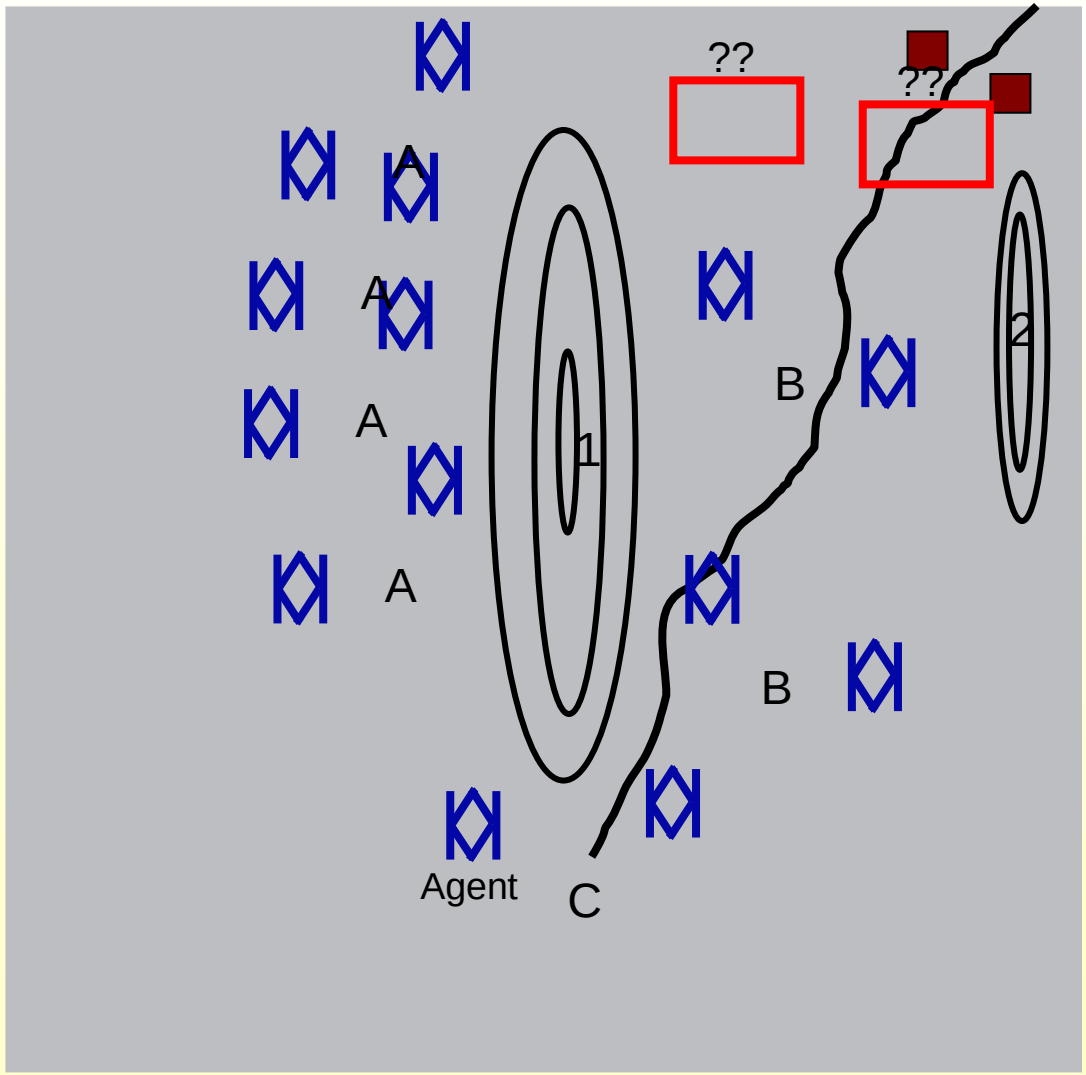
COGNITIVE CAPABILITY VISUAL SIMULATION EXAMPLE

- **Reconstruct image of last known situation**
- **Visually simulate time based on domain temporal/spatial factors (vector)**
- **Add relevant objects**
- **Apply domain knowledge**
- **"Sense"**

  Where are my elements? Where are they in relation to each other? To the terrain (hill 1, hill 2, road)?

  Where is the enemy's center? Where are my forces in relation to the enemy?

  How much time for A to move to a position north of hill 1 oriented east? For C to move up to B's position?



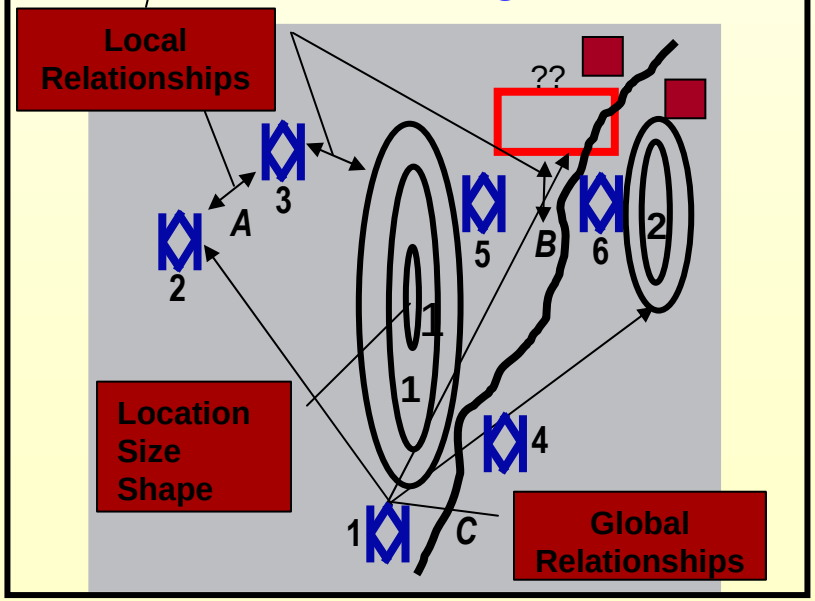
SUMMARY

Symbolic (Descriptive)	Visual Image (Depictive)
Explicit, abstract objects	Implicit, concrete, objects
Explicit relations	Implicit relations
Location, size, shape, features, orientation optional	Location, size, shape, features orientation matter
Computationally efficient for maintaining local relationships	Computationally efficient for maintaining global relationships

Symbolic

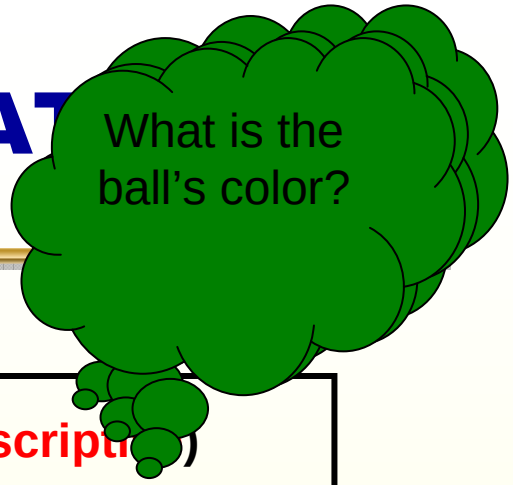
Object (scout-1) Object(scout-4)
Composite-Object (Section-C)
Contains (Section-C, scout-1, scout-4)
Object (hill-1) Object(hill-2)
Northeast (scout-3, scout-2)
North-of (enemy, section-B)

Visual Image



*Adapted From Kosslyn Image and Mind

BACKGROUND & MOTIVAT



Symbolic (Descriptive)	Image (Depictive)
Explicit, abstract objects	Implicit, concrete, objects
Explicit relations	Implicit relations
Size, shape, orientation optional	Size, shape, orientation matter
Object features optional	Object features - Color - Texture - Part
Computationally efficient for maintaining local relationships	Computationally efficient for maintaining global relationships

Symbolic (Descripti)

Ball

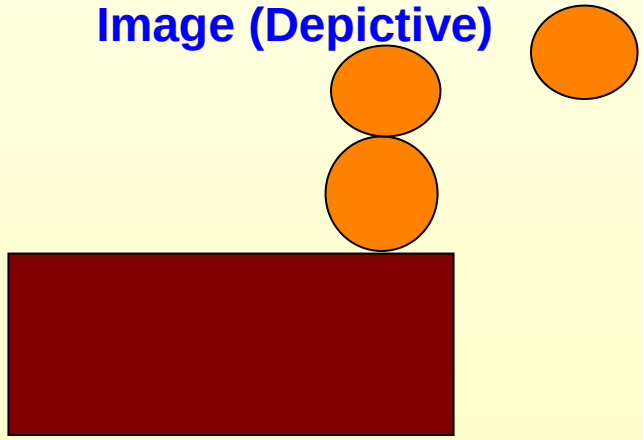
Box

On (Ball, Box)

Right-Side(Ball, Box)

Orange(Ball)

Image (Depictive)



*Adapted From Kosslyn Image and Mind

RELATED WORK

- Psychology
  - Cognitive Scientists/Vision Scientists
  - Mental Imagery Debate

- Neurology
  - Temporal Cortex (What - Visual)
  - Parietal (Where - Spatial)
  - Occipital (What + Where)

- AI
  - Diagrammatic Reasoning
  - Spatial Reasoning
  - Computational Imagery Models
  - Robotics

- Applications: Simulation, Robotics, Decision Support Tools, Intelligent GIS

COGNITIVE CAPABILITIES

- **Reasoning**
  - Anticipating
  - Remembering
  - Emergent Objects & Relationships
- **Planning**
  - Visual simulation
  - Time/distance analysis
- **Learning**
  - Mnemonic
  - Identifying new visual features relationships or not previously made concrete
- **Motor Skill Acquisition and Rehearsal**

COGNITIVE CAPABILITIES

- **SPATIAL REASONING**
  - Connected/disconnected
  - Part of
  - Inside/Outside
  - Enclosed/Not Enclosed
  - Relative spatial position
    - Preposition-based (above/below, left-of/right-of, adjacent)
    - Compass-based (north-of, east-of, etc)
  - Relative size
  - relative distance (near/far)
- **SURFACE PROPERTY REASONING**
  - color
  - shape
  - texture
  - contrast
  - presence of an object's part

*Mental imagery used when not found in symbolic information

* "Just-in-time" (JIT) reasoning

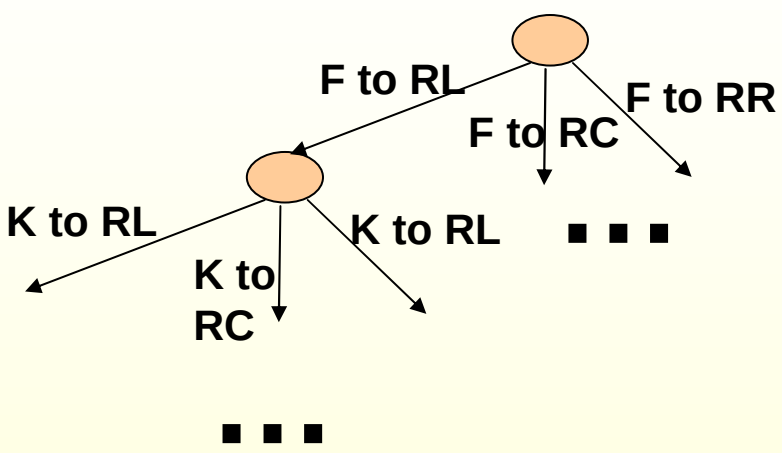
PSYCHOLOGICAL CONSTRAINTS (1 of 2)

- **General**
  - Integrated into cognitive architecture
  - Perform in real time
  - Chunk over results
- **Encoding**
  - Multi-modes (metric and symbolic)
  - Symbolic relationships used to coordinate juxtaposition of individual parts
  - Image primitives are region based and/or point based
- **Storage**
  - Store encodings in one of the Soar memory elements
  - Initial Research: Working memory contains indexes, “Image Memory” contains geometric data
  - Future Research:
    - Image primitives in episodic memory?
    - Conceptual, symbolic relationships in semantic memory?

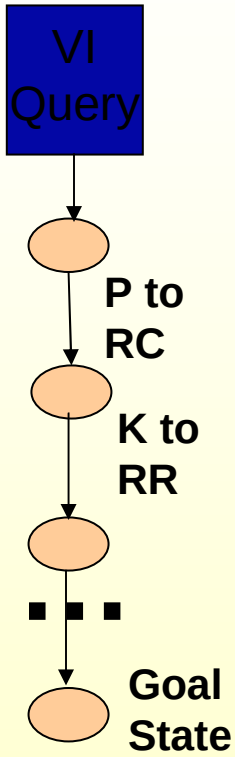
PSYCHOLOGICAL CONSTRAINTS (2 of 2)

- **Retrieval**
  - Generated from separate retrievals
  - Images constructed by examining material already in visual buffer
  - Object parts retrieved on an “as-needed” basis.
- **Display (Visual Short Term Memory)**
  - Fixed spatial extent (fixed real estate)
  - Roughly circular near center, elliptical on outsides
  - Resolution is highest in the center of the image and decrease as you move towards the periphery of the image.
  - Image fades over time
  - Default size of images in display based on
    - Relative size of object being generated
    - Nature of material
    - Problem space
  - Default orientation is Egocentric? (Self as origin)
- **Use**
  - Capable of spatial and surface property reasoning
  - Transformations (zoom, scan, rotate, etc) support reasoning and occur in a stepwise fashion.
  - Must allow generation of new representations (learning)

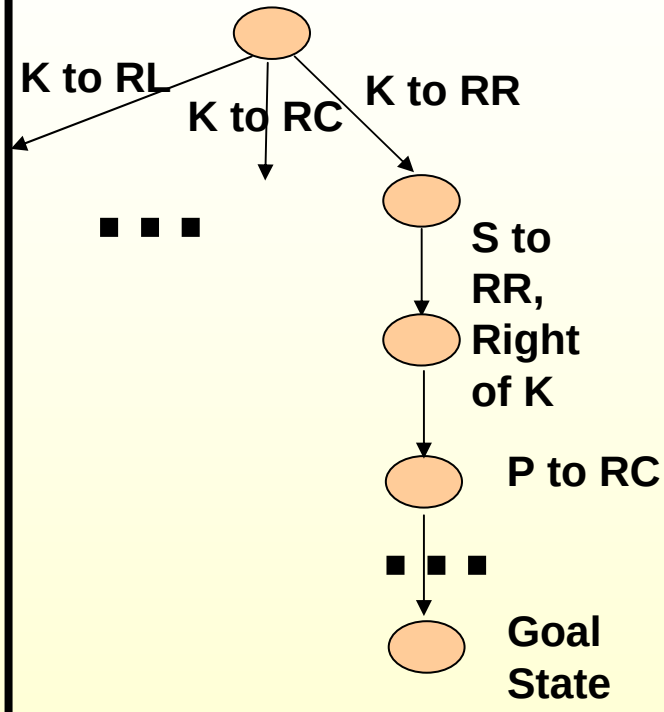
AGENTS
PROBLEM SPACE SEARCH



Agent Random



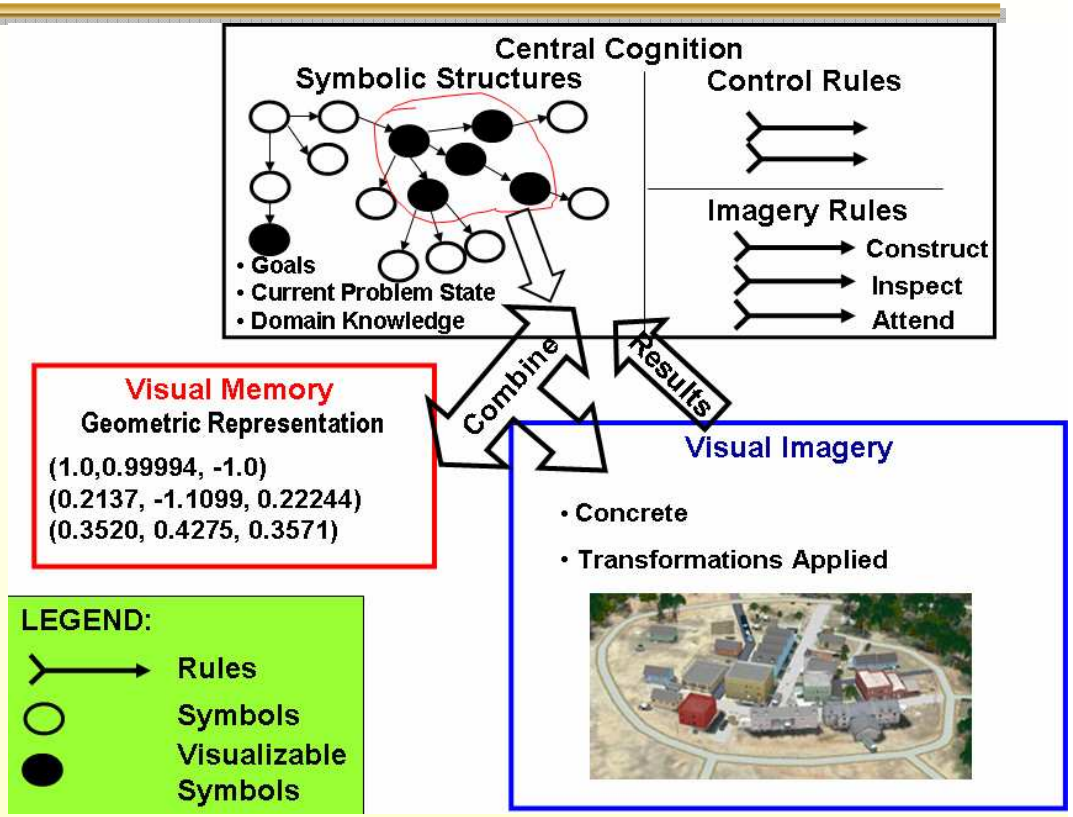
Agent with Imagery



Agent using Least Constraining Value (LCV)

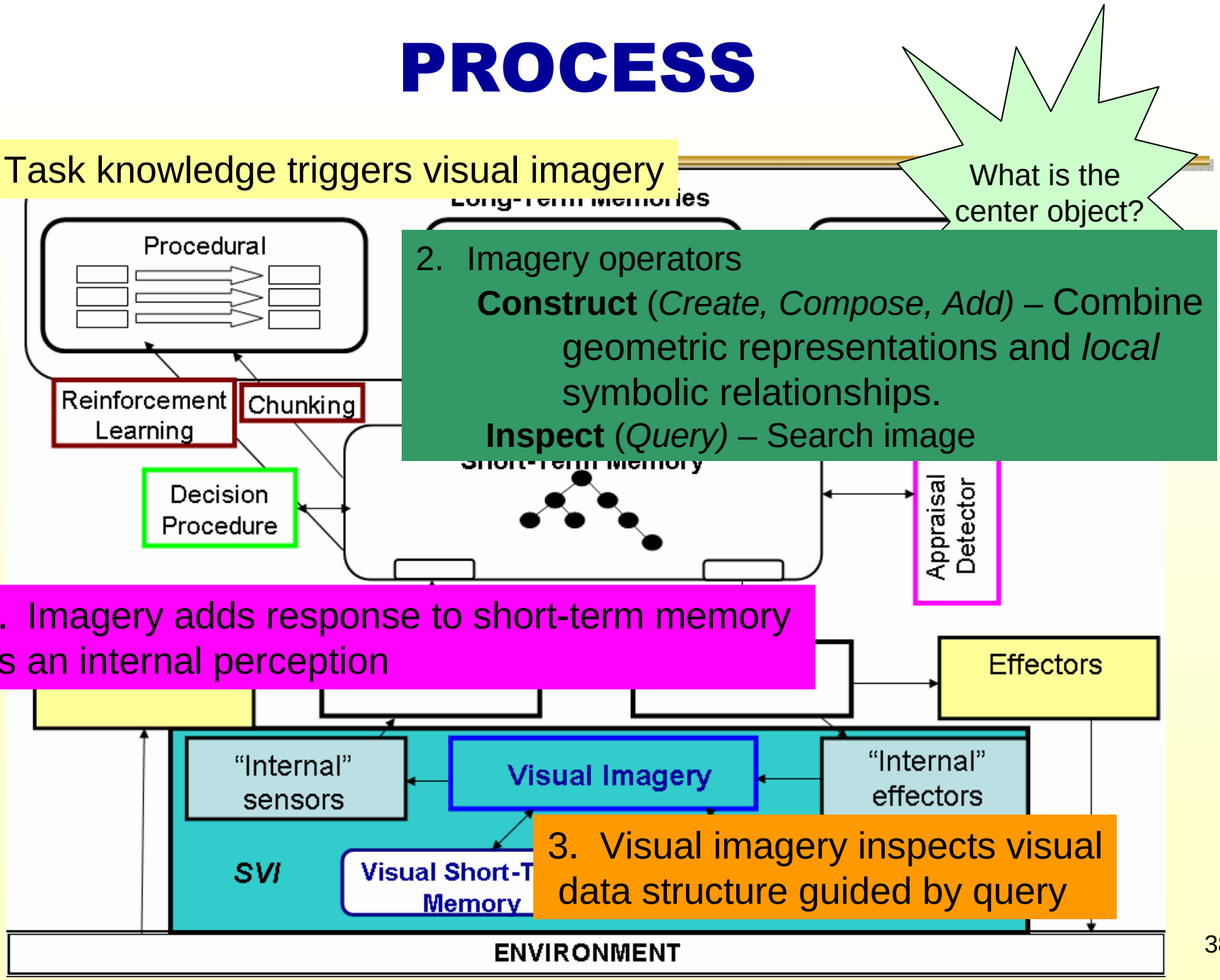
THEORY

- **Hypotheses**
  - – Symbolic representations are sufficient but are not the most efficient.
  - –Visual imagery is a subsystem of the total cognitive architecture
  - – Visual imagery shares mechanisms with vision
  - –The composition of an object and local relationships stored in central cognition's memory structures
  - – An object's metric representation (shape, color, texture) stored in visual long-term memory.



PROCESS

1. Task knowledge triggers visual imagery

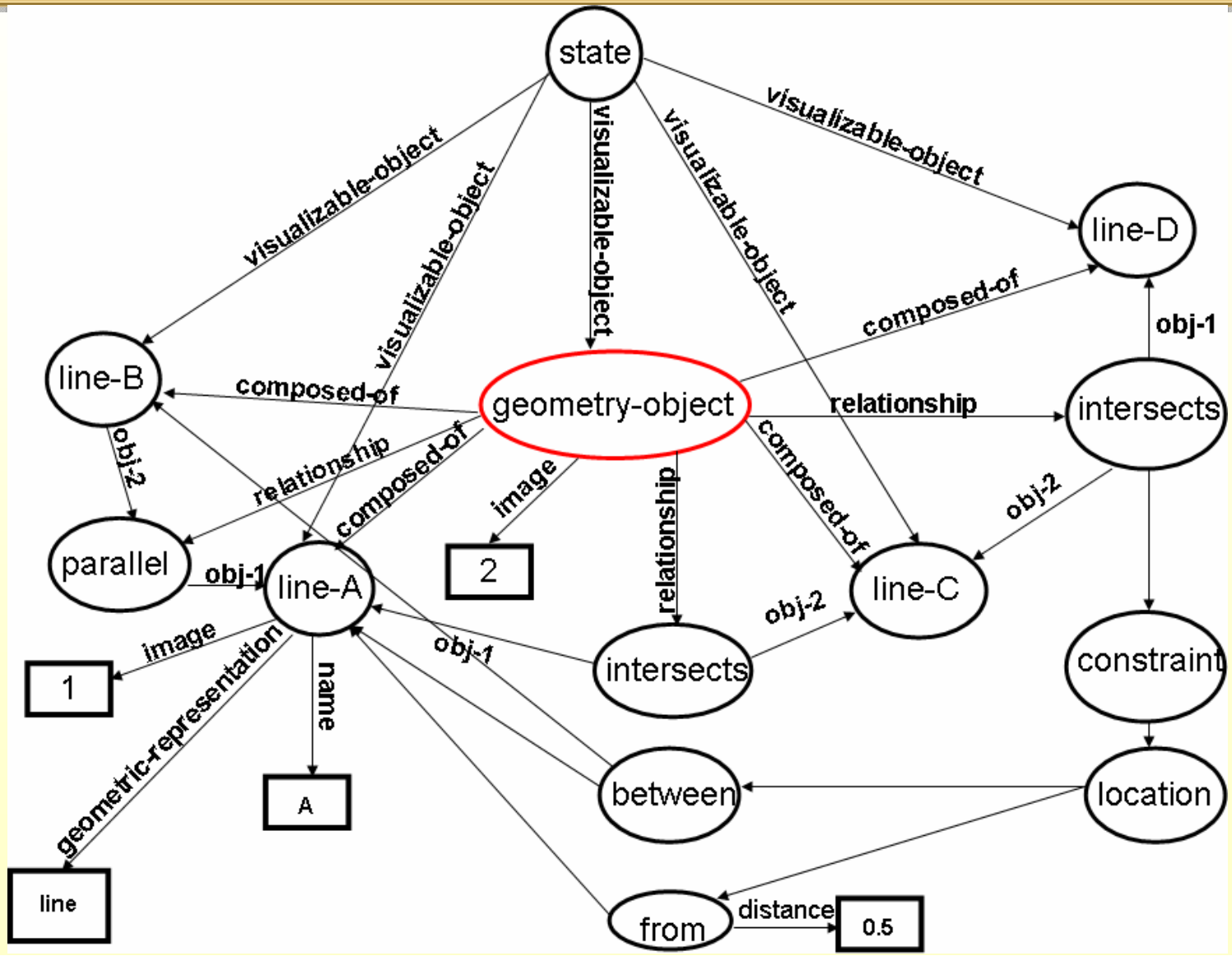


2. Imagery operators
   Construct (*Create, Compose, Add)* – Combine geometric representations and *local* symbolic relationships.
   Inspect (*Query)* – Search image

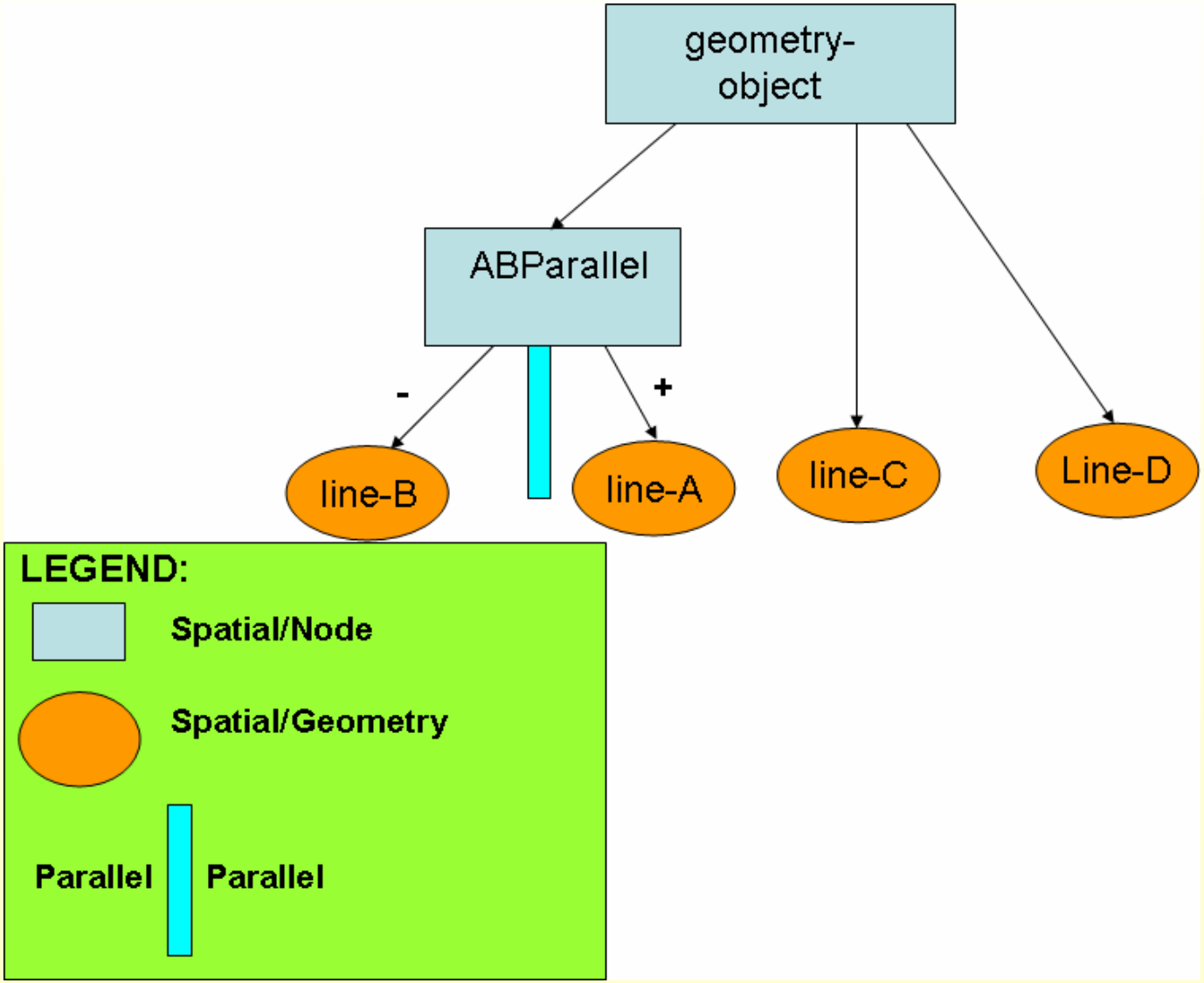
3. Visual imagery inspects visual data structure guided by query

4. Imagery adds response to short-term memory as an internal perception

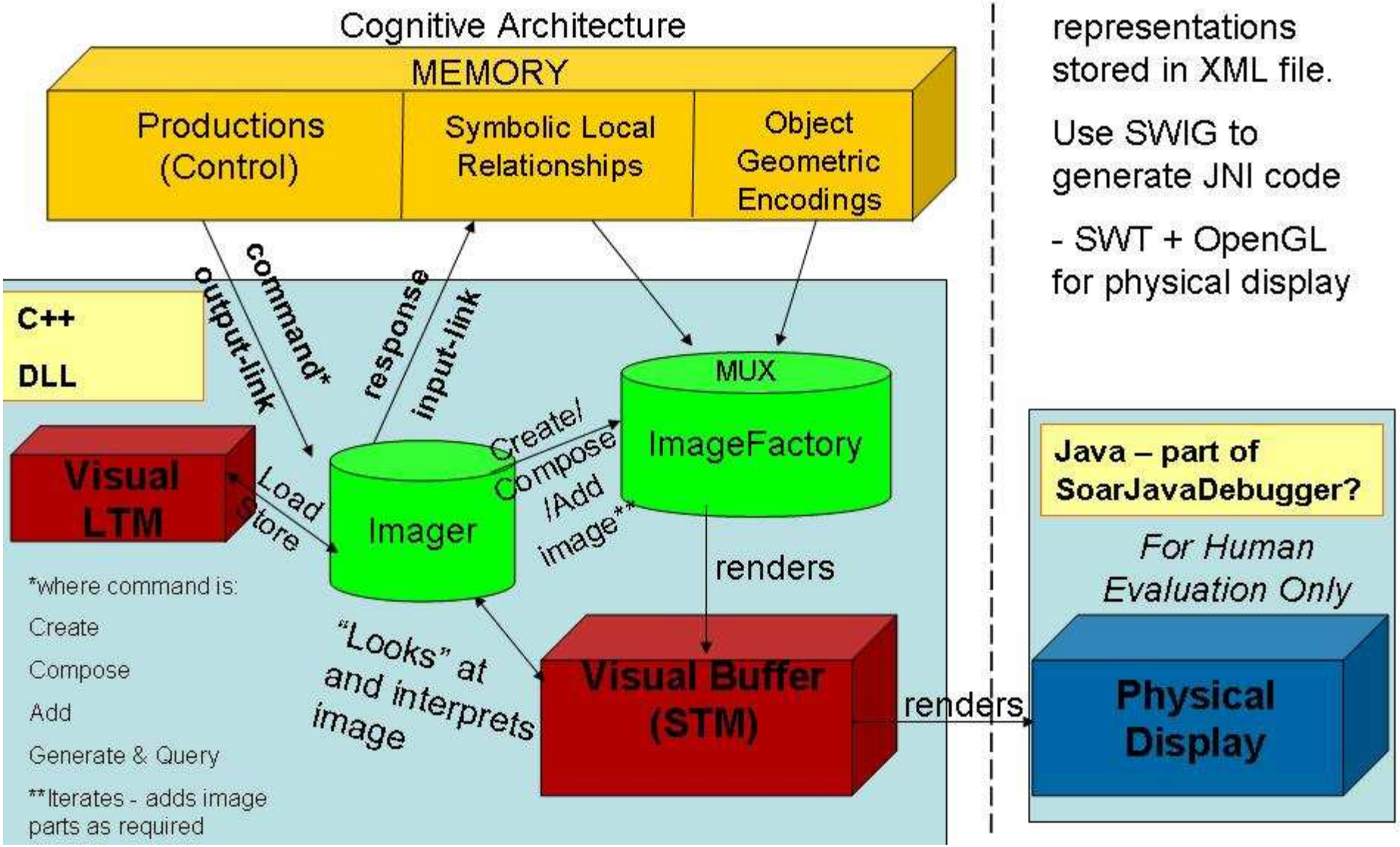
SYMBOLIC REPRESENTATION



VISUAL REPRESENTATION



STRUCTURES



Notes:

- Object Geometric representations stored in XML file.

Use SWIG to generate JNI code

- SWT + OpenGL for physical display

STRUCTURE
SCENE GRAPHS

- All nodes in DAG are Spatial Objects
  - Represents a coordinate system in space
  - Transformation Data
  - World Bound Data
  - Effects, lights, and global state data
- Internal Node Objects
  - Grouping responsibility
  - Parent World Transformation data
  - Parent World Bound data
- Geometric Objects
  - Leaf nodes
  - Model vertices and indices
  - Model normal
  - Model scale
  - Model bound
  - Only need to keep 1 instance of an object's geometrical representation (share across images)

Egocentric-to-Allocentric Transformation

- Egocentric (viewer centric):
  - Standing over the sink
  - Tub behind me
- Allocentric (viewer neutral):
  - Torch above tub
  - Doorway south of tub
- Hippocampus known to be responsible for egocentric → allocentric transformations
- Task: answer ?s about non-viewer centered relationships in a familiar scene (incomplete specification)Soar lacks any intrinsic mechanisms for spatial perception and reasoning

Spatial Reasoning vs. Criteria

- Emphasize gaps of one approach
  - Soar doesn't do this; SoCs don't do this?
- Require end-to-end performance
  - Yes.
- Tap multiple functional capabilities
  - Yes (episodic memory, deliberate reasoning, perception)
- Require learning
  - Yes (episodic learning)
- Have human data for comparison (psych, neural)
  - Some data has been collected for this kind of task ("Jon")
- Real-world or laboratory tasks? (eg, task switching) scaled-down simulations of real-world tasks (Eaters).
  - Laboratory but has real-world flavor; sub-component of decathlon challenge
- Across a range of time-scales
  - Yes (individual events/experiences recurring across longer time scales)
- Number of trials needed for learning?
  - Yes. Must gain familiarity with environment in order to be able to answer.

PROCESSES

- **Create***
- **Compose***
- **Add object to image***
- **Generate and Query***
- **Transformations to support mental imagery processes**
  - **Translation/Camera changes: "Scanning"**
  - **Zoom in/out: "Scaling"**
  - **Rotation: "Mental Rotations"**
  - **Compositional Transformations: "Mental Simulation"**
- **Object Feature Queries**
  - **Color, Texture, Shape**
- **"Animation" to support mental simulation.**

*Implemented

PROBLEM DOMAIN

Region Left	Region Center	Region Right
? **?** 		

Start State: Each object is on the counter.

Operators:

(1) Pick up and move an object to a region

(2) Move an object within the region.

Goal State:

Region Left: Fork above Napkin

Region Center: Plate

Region Right: Knife left of spoon

PROBLEM DOMAIN

Region-Left	Region-Center	Region-Right
		

What is the center object?

FUNCTIONAL CAPABILITY

- "Just in time" global spatial relationships
  - Do not have to encode all global relationships
  - Can query imagery as required
  - Center object not inferable with symbols and initial knowledge
- Other potential functionality
  - Size comparisons (what is the largest object in scene)
  - Visual features (color, texture)
- Taking advantage of close coupling with vision

DEMO