

# Multi-modal Representations as the Basis of Cognitive Architecture

B. Chandrasekaran

Laboratory for AI Research

The Ohio State University

# “Traditional” AI

- “Thought” is something like a sentence in natural language
  - A proposition
    - “Cat is on the mat,” “I am holding a cup of coffee.”
  - Or, an “attitude” towards a proposition
    - I want it to be true that “I am holding a cup of coffee.”
- “Thinking” is having a series of thoughts, generally in pursuit of a goal (which is an attitude to a proposition). (Language of Thought Hypothesis)
- Thinking is achieved by manipulation of a representation of the world and our knowledge of the world, all of which are represented as propositions.

# Steps in the Argument

- A view of Reasoning w. external Diagrams
  - Representations span the agent and the environment, and jointly each state is *bi-modal*.
- Make diagrams internal.
  - Problem solving with mental images of diagrams. Now internal state is bi-modal
- Generalize to multi-modal state:
  - External: [External representations in general: e.g., designer with a clay model, chemist with a 3-d molecular model, a composer with a musical instrument in hand]
  - Internal: No reason to restrict the notion of an image to the visual
- Outline a multi-modal cognitive state and elements of the associated engine.
  - How the problem-space in a Soar-like architecture may be multi-modal
- Discuss why this kind of architecture is useful for any agent, natural or artificial.
- Issues in implementing the multi-modal state:
  - Specifically, what kind of rep. frameworks support representing images in such a way that they are both image-like and symbol-like?

# Use of Perceptual Representations in Problem Solving

- In “diagrammatic reasoning,” a problem solver uses a diagram as part of a PS episode.
  - Proposition extraction.
    - Proposition extraction is direct. Often the same proposition might take a chain of inferences if done within a propositional representation.
  - Deliberative reasoning with visually extracted & other propositions and rules of inference.
  - Proposition projection, “simulation” of motion and changes in positions result in new external representations from which additional propositions may be extracted.

# Perceptual Reasoning Uses All Processes Opportunistically

- Projection and simulation may enable extraction of new propositions
  - Which in turn may enable additional propositions to be inferred using conceptual knowledge, which in turn may be projected,...
  - Very important to understand how projection & simulation make new propositions available for extraction

# Reasoning (Inference-Making) (contd)

- Propositions extracted from perceptual representations often replace computationally more expensive inference chains.
  - Defining the conditions under which the the propositions that are extracted correspond to correct inferences in the general case is an interesting issue that need not concern us here.

# Problem state is bi-modal

- Problem representation spans – extends over -- the cognition of the problem solver and the environment.
- Together, the representation is bi-modal.

# Internal Visual Representations (Images)

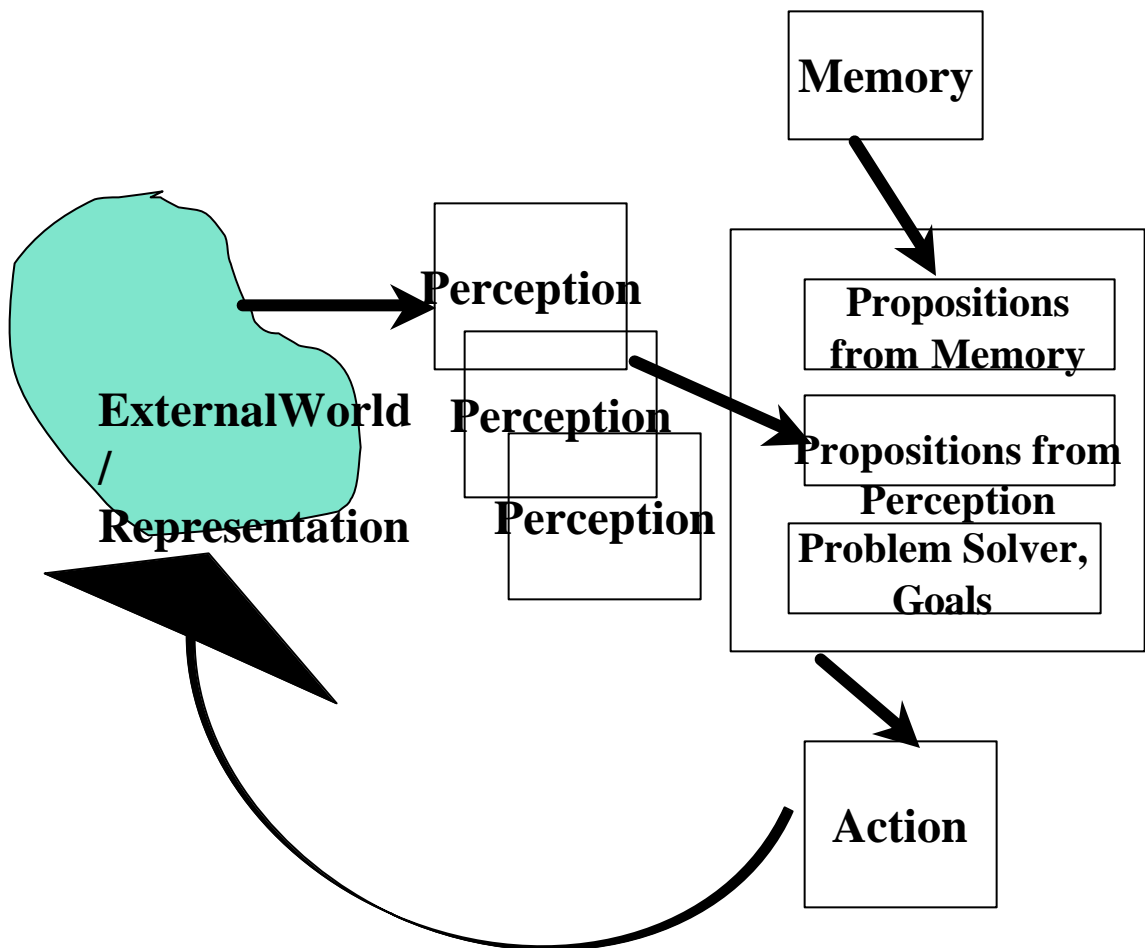
- Controversy about mental images notwithstanding, in many PS situations, mental images of diagrams are used in problem solving, playing essentially the same *functional role* as the external diagrams do.
  - Examples.
- In a real sense, the internal representation is bi-modal.
- Question: what kind of internal representation frameworks explain the functional role as diagrams without being diagrams to be perceived?
  - I'll hint at a solution.

# Thinking: The Canonical View

- Thinking is a process of manipulating a representation of the world of interest
- The representation is something like NL sentences, LOT hypothesis
- Thoughts are propositions (or attitudes to them, such as believe P, desire P, etc.)
  - From “thoughts have propositional content,” to “thoughts are representations of propositions.”
- This general stance towards cognition is not specific to “logic” approaches in AI. Approaches based on frames, scripts, rules are all in this sense “symbolic AI.”
- For our purposes, connectionism and related ideas are alternative ways of representing the same type of information, and thus orthogonal to the issues raised here.

# Interaction with the External World in the “Standard” View

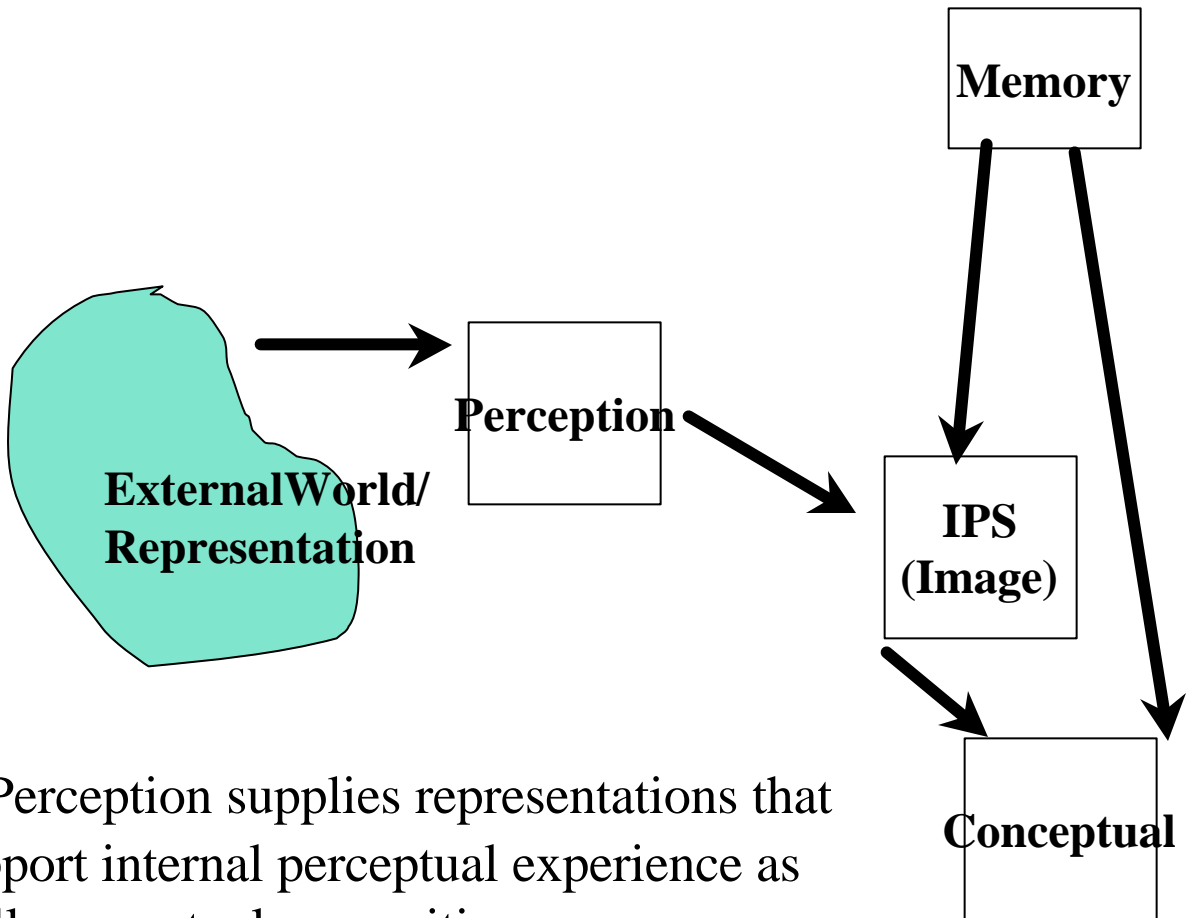
- Interaction with the world takes place by taking in knowledge of the world from perception in the form of propositions and generating propositions about actions to take that are then executed by motor systems.



# What is really in a Cognitive State?

- OTOH...
- Phenomenologically, if we analyze the content of thought, we are aware not only of elements that have a *propositional* content, but also elements with *perceptual* content.
  - We can “see” a child swinging in the yard, we can “hear” tunes and have difficulty getting rid of them. We can decide if we can go through a passage by imagining if we can contort our bodies to the way required.
  - Even in communication, we may use, in addition to language, pictures, 3-d models, gestures, music, and so on. Thus, there is no good reason to model inner thinking purely on language and its structure.

# Perception Supports Both an Inner Perceptual Experience and Conceptual Knowledge



- Perception supplies representations that support internal perceptual experience as well conceptual propositions
- Inner perceptual experience can also be created by representations from memory, and of course memory can also supply conceptual propositions as well.

# Images are Not Just For the Visual Modality

- Much of the work (debates) in CogSci on images has been in the visual domain, but, as the example of tunes suggests, the phenomenology of images is not restricted to the visual.
  - “I can almost taste the food.” In addition to all the perceptual modality, one can have kinesthetic images as well
  - A sense of the contortions of the body when imagining going through a narrow passage. On looking at a design diagram: “The mouse buttons feel like they are too far apart for comfort.”

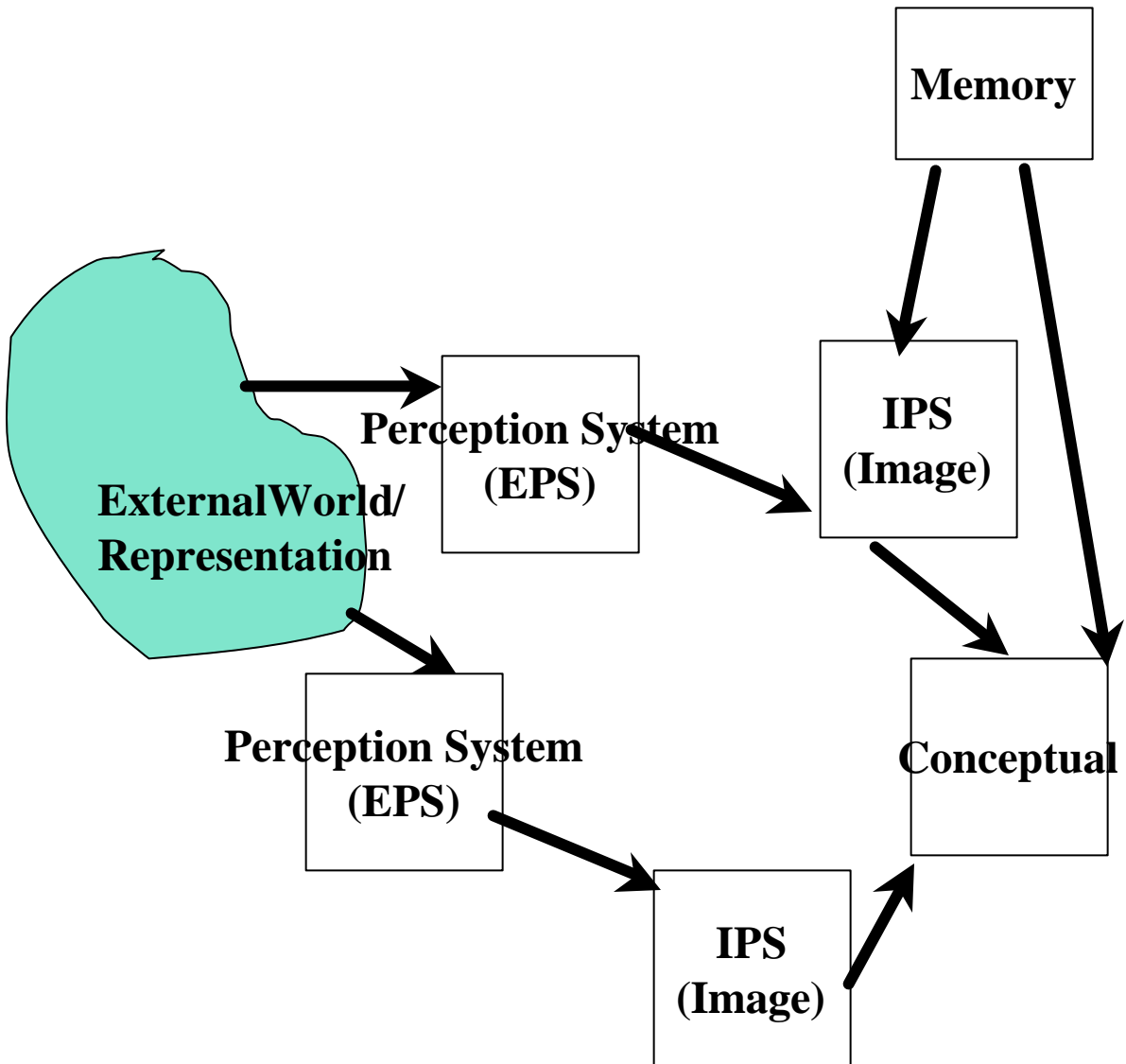
# Images are not identical to their propositional content

- Two dimensions to images:
  - The first is the experiential dimension
    - Listening to music or looking at a sunset is not identical to being given propositions about them.
  - The second is the informational dimension
    - An image corresponds potentially to an infinite number of propositions.
    - Inferential or information extraction operations are different for images.

# Internal Multimodal Representations: A Proposed Functional Architecture

- Each state of thinking is *potentially* multimodal.
  - Potentially, because not all instances have all modalities.
- Perceptual modalities (PM) includes kinesthetic modality for this discussion

# More than One Perceptual Modality



- Each modality supplies both an image that is experienced as well conceptual predicates
- Remember that kinesthetic modality is one of the perceptual modalities in the above.

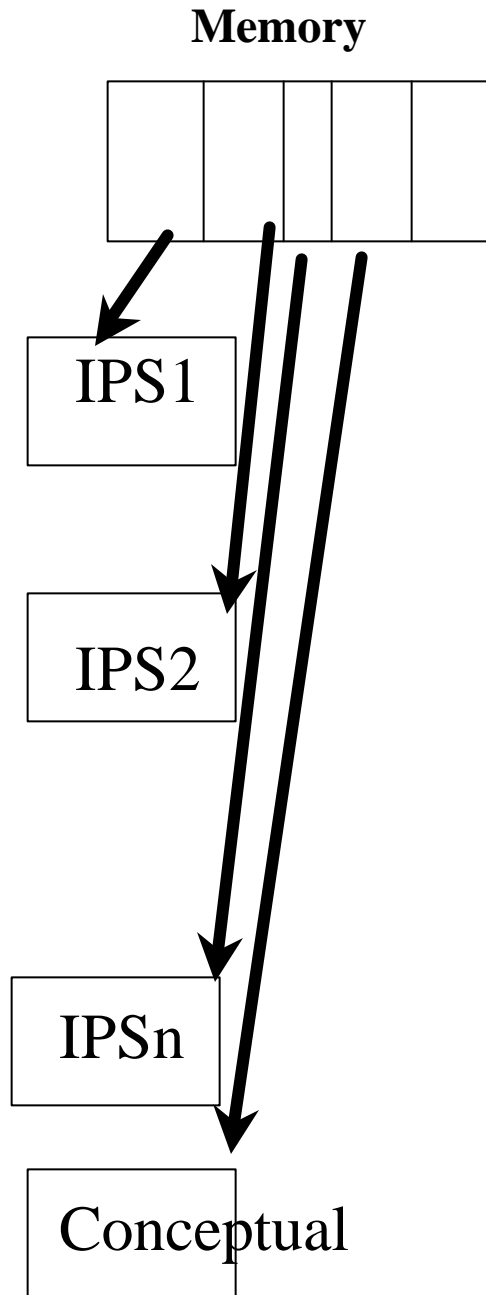
# Functional Architecture: Multimodality of Cognitive State

- The effects of acts of perception and of imagining are very similar, except that EPS can maintain the richness of the images in IPS without strain on memory.
- Awareness (cognitive state) is multimodal, as a rule.
  - Its components are the various IPSs and the *conceptual* modality.
    - For the sake of uniformity, we refer to representations in IPS's as well as the conceptual modality as images. And when we say IPS, we will include the conceptual mode as well.

# Functional Architecture: Memory is multimodal too

- Agent's memory is also multimodal, paralleling the organization of the cognitive state
  - Views, postures, tunes, concepts, episodes that have all these..
  - Elements in one memory mode are associated in various ways with elements in other modes
    - Concept of apple in memory may be associated with the memory of its shape and color in the visual modality, the act of biting into it in the kinesthetic memory and so on

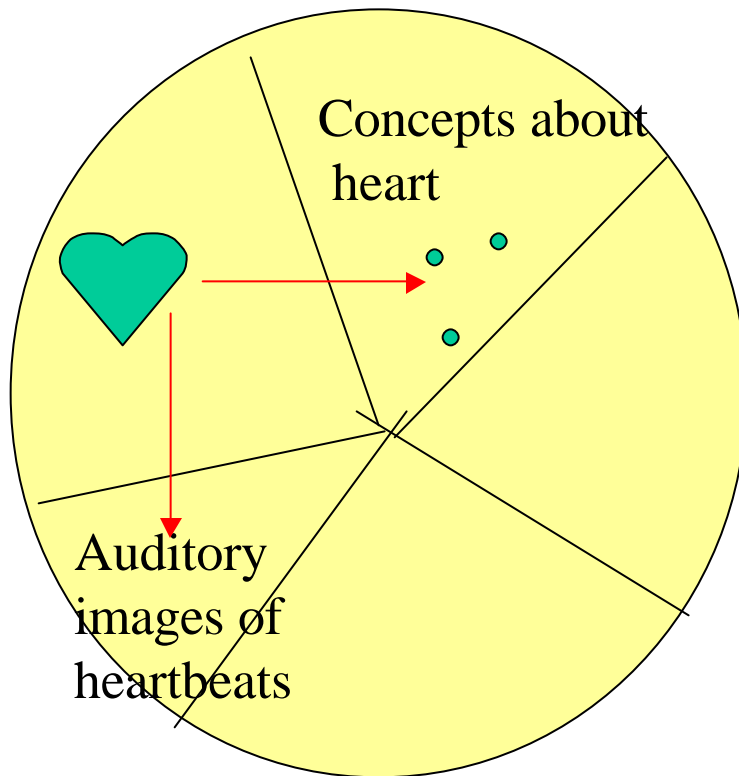
# Memory is Multimodal Too



## Representations May Evoke Associated Representations

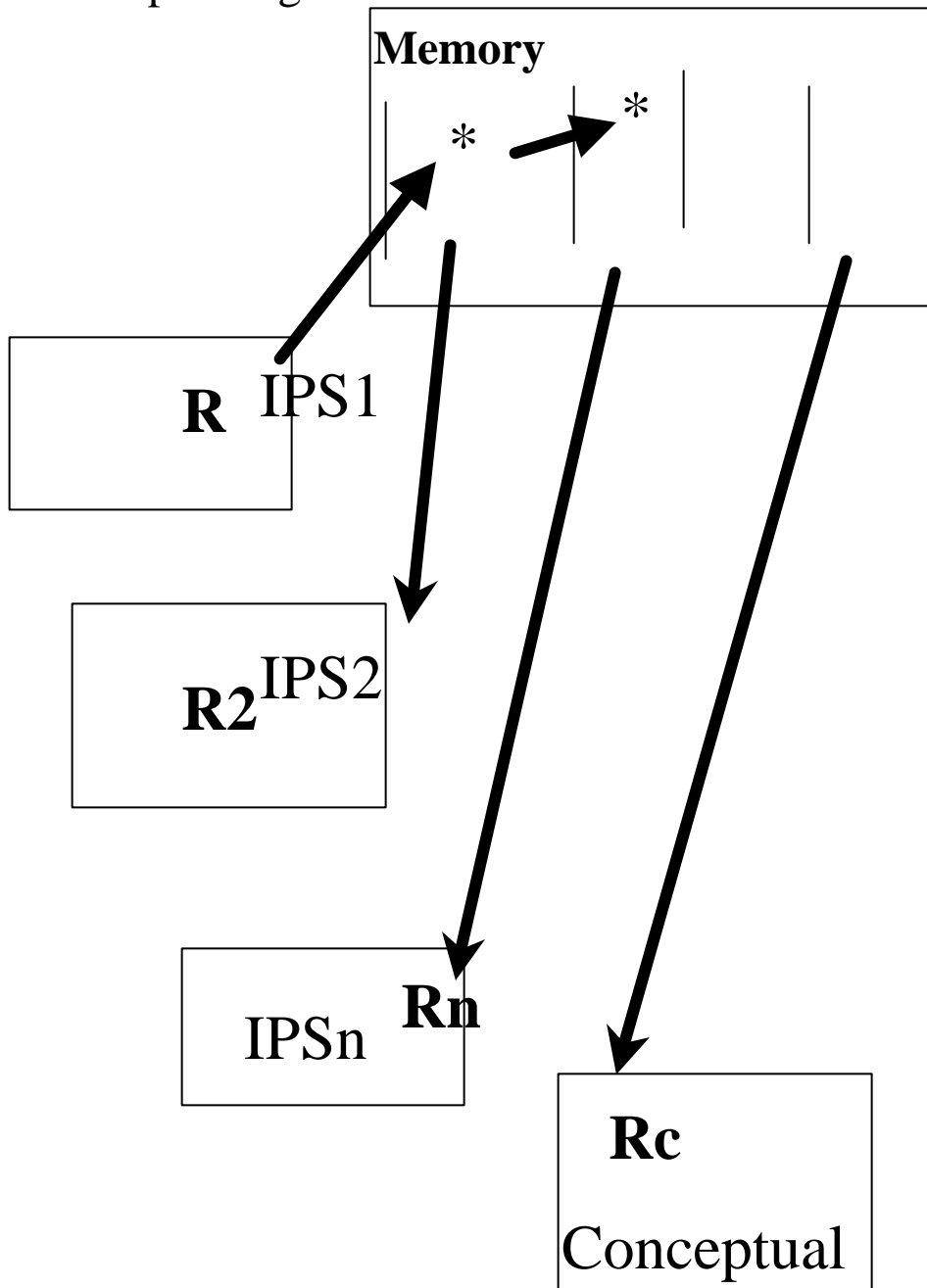
- A representation in any IPS (I.E, in the cognitive state) may evoke images in other IPSs
  - The evoked images are those associated in memory
  - E.G., The visual image of a heart may evoke conceptual information about its role in life and health issues, the auditory image of heartbeats,.....
  - These multimodal evocations occur whether or not the image in an IPS came from EPS or from memory

# Intermodal Evocations

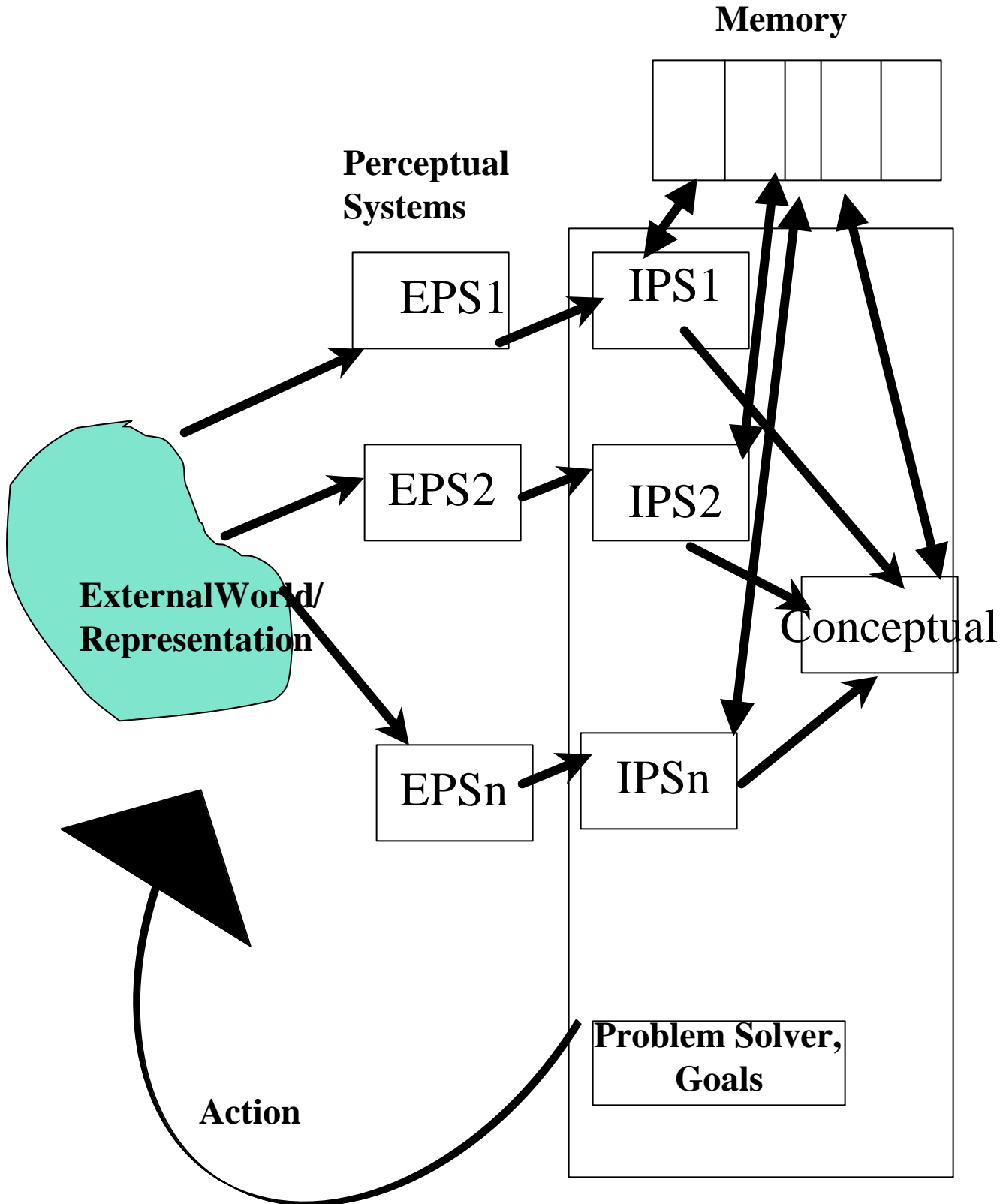


# Intermodal Evocations

A representation in modality may evoke associated representations in memory of other modalities, that in turn would evoke images in the corresponding IPSs.



# Cognitive state of agent is multimodal



# The Architecture is a Functional One

- I am only appealing to the *capabilities* or *functionalities* of imagining and perceiving and reasoning.
  - Alternative implementations of these capabilities in different mechanisms (Turing symbolic, connectionist, dynamical systems, whatever) possible.

# Why This is Not Just Another Instance of a “Standard” Propositional Representation

- The symbols and relations in IPS do of course refer to objects in some world, but relations between them is not abstracted into a relational symbol.
  - Each image potentially corresponds to an infinity of propositions.
  - The set of operators for each modality is an “analog” of the corresponding modality.

# What do agents do with IPS representations?

- Each of the IPS components supports its own characteristic form of inference
  - Perceptual modalities support perceptual proposition extractions, while the conceptual modality supports “reasoning”
- Intermodal evocations provide a powerful way to bring to bear distal knowledge in all modalities to bear on the current situation
  - Shape & color of apple (visual) --> taste of apple (taste) --> appropriateness for the pie recipe (conceptual) --> decision to buy apple (conceptual)

# IPS Representations Drive Problem Solving

- Problem solving is a process in which the agent's cognitive state changes as a function of the contents of the current cognitive state, the state of the external world, and the PS goals
- Changes in the external world (or attention) cause EPS to deliver new percepts and relations to IPS and the conceptual component
  - Proposition projection is one cause of the change in the external world
- Changes in one IPS representation may evoke associated images in other IPSs
- New propositions may be extracted in those IPSs, including in the conceptual mode
- Conceptual inferences may include action items that change the external world
- Goals determine control of which inference options are pursued

# Advantages to the Agent

- A wide-variety of modality-specific information extraction operators are directly available -- obviating the need for complex inferences from propositional abstractions.
- Each image corresponds literally to an infinity of propositions. Thus, is experience is stored closer to perception -- how it was experienced -- propositions can be extracted as appropriate for the task at hand.
- Builds on top of perceptual machinery already needed for other purposes.
- Continuity with animal intelligence in general.

# Connecting to Other Cognitive Architectures

- Let us use Soar as an example
  - All long term memory in the form of productions. Each production is a condition-action pair. The conditions form access paths and the actions form the memory contents. Productions continuously match against a declarative working memory that contains the momentary task context.
  - SOAR achieves all cognition by *search in problem spaces*, and architecturally supports this by a flexible, two-level *recognize-decide-act* control structure.
  - Actions consist of applying *operators* that change problem states. (Problem spaces and operators arise from the productions that are relevant to the current goal.)
  - *Impasses* set up subgoals.

## Soar (continued)

- Soar's problem space consists of problem states, described in terms of predicates that describe the state.
  - The proposal is to augment Soar's problem state description to be multi-modal.
  - Different operators can be available for different modalities (information extraction operators for perception/kinesthetics, inference operators for the conceptual part).

# Imagery Debate & Issues

- Image versus propositions (old debate, dated now)
- Experience of imaging is strong, plus experiencing of applying certain visual operations is strong as well.
- OTOH, propositionalists said:
  - Imagery not stored as a picture; needs a homunculus, postpones problems, not solve them.
  - Composability of images suggests a symbol-structure.
  - You don't normally perceive an image: all its constituents already interpreted
  - Nonuniformity of detail in imaging: pictures (unless torn off or smudged) do not have this.
  - Ambiguous figures: if you form one percept, and imagine, hard to form the other percept.
  - Vague, indeterminate images (abstract hat, elephant, etc.)
  - Thus, it is all propositions

# Issues in the Implementation/Realization of IPS's

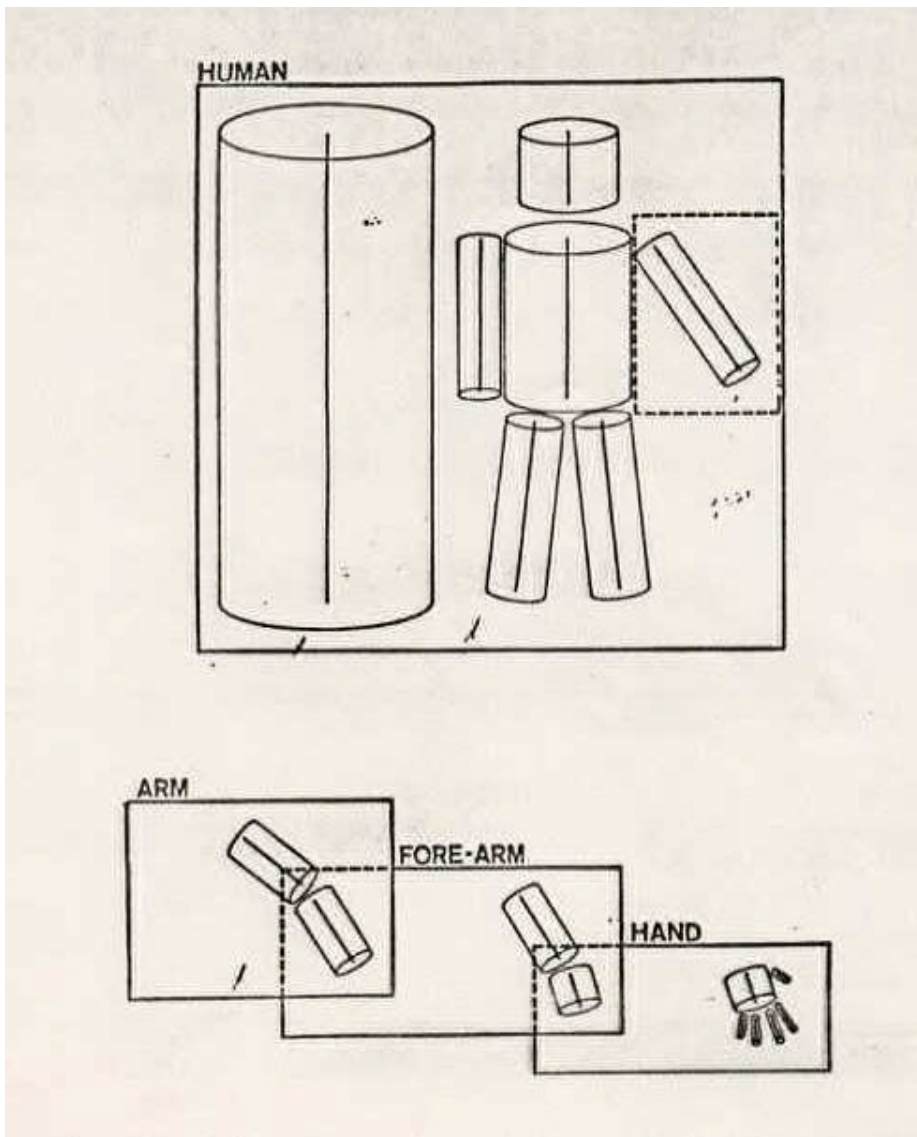
- The solution has to respond to both sets of concerns, i.e., we need to see how it has all the image properties we need and the symbolic properties we need as well.

# A Brief Outline of the Solution

- Outline of solution:
  - Let us simplify, for pedagogical purposes, that the output of visual perception is a 3-d description of the shapes of objects and their relations in the scene.
    - (This is pre-object-recognition/naming, kind of what you see when you see a Henry Moore sculpture.)
  - Step 1: Choose a vocabulary of shape primitives and relations.

# 3-D Shape Representations

- Example
  - Marr-Nishihara generalized cylinders



# 3-D Shape Representation (contd)

- Other examples: Binford generalized cones, Biederman shape primitives
- New possibilities: Fourier- or wavelet-like infinite series; singularity-based shape decompositions,...
- They all have the following properties
  - Numerically parametrized primitive shapes and relations
  - The target shape can be represented as a hierarchical symbol structure, the ISS.
- (Similar stories for other perceptual modalities.)

# A Brief Outline of the Solution (contd)

- Step 2: Output of perception is a tree structure of primitives and relations (the **ISS**). Top levels of the tree give approximations of gross shapes, while additional levels add levels of detail.
- Step 3: What is stored in memory is just this numeric-parametrized symbol structure. This structure is only meaningful in the visual modality.
- Step 4: When this structure is retrieved from memory and the placed in the WM corresponding to the visual modality, the effect is the same as if this structure is the output of perception.

# ISS can provide a framework to resolve the paradox

- What is stored in memory are subtrees of the ISS. The aggregation of primitives corresponds to whatever percepts are active.
- Criteria for selection of subtrees:
  - Chunk size limits
  - Abstraction principle: OTBE, top levels get retained, details lost (accounts for vagueness of form)
  - Attention (interest, goals..) selects subtree

# Perceiving & Imagining

- Having a perceptual experience is having a corresponding ISS. Thus, in imagining we retrieve/construct ISS's which then give rise to the sense of the perceptual experience
  - Kosslyn: Significant overlap in excited regions in the visual cortex between perceiving and imagining
  - Konorski: targeting reflexes are active during seeing and dreaming, but not during imagining.

# What does this mean for AI

- Especially important in the emerging integrated systems approach to AI -- robotic-based intelligence systems, with perception, action and reasoning rolled into one.
- As the robot experiences the world, its memory and reasoning exploit the structure of perception.

# Concluding Remarks

- Experiencing, problem solving, reasoning, takes place in the context of an external world that we perceive, act on, imagine and reason about in multiple modalities
  - Conceptual modality is always present
  - Different tasks emphasize different perceptual modalities
    - Music composition vs mechanical design
- In our framework, the internal representational life of the agent is multimodal, with representations in one mode evoking allied representations in other modes, and each mode making inferential contributions for which it is best suited. Mental images come in all modalities.

## Concluding Remarks (cont.)

- Treats conceptual component as just another component with equal status with inner perceptual and kinesthetic components. In one way of thinking, having concepts is imaging the world in the conceptual world, just as having images is imaging the world in the perceptual mode.
- Logical rules of inference are just a very small part of the information extraction operators in the conceptual part.