Multimodal Representations as the Basis of Cognitive Architecture

B. Chandrasekaran

Laboratory for AI Research

The Ohio State University

Steps in the Argument

- A view of Diagrammatic Reasoning (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.
- Generalize to multi-modal state: No reason to restrict the notion of an image to the visual domain.
- Outline a multi-modal cognitive state and elements of the associated engine. Compatible with the problem state notion in Soar, though details remain to be worked out.
- Discuss why this kind of architecture is useful for any agent, natural or artificial.

Use of Perceptual Representations in Problem Solving

- In "diagrammatic reasoning," a problem solver uses a diagram as part of a PS episode.
  - Proposition extraction and projection.
  - Deliberative reasoning with visually extracted & other propositions and rules of inference.
  - "Simulation" of motion and changes in positions result in new external representations from which additional propositions may be extracted.
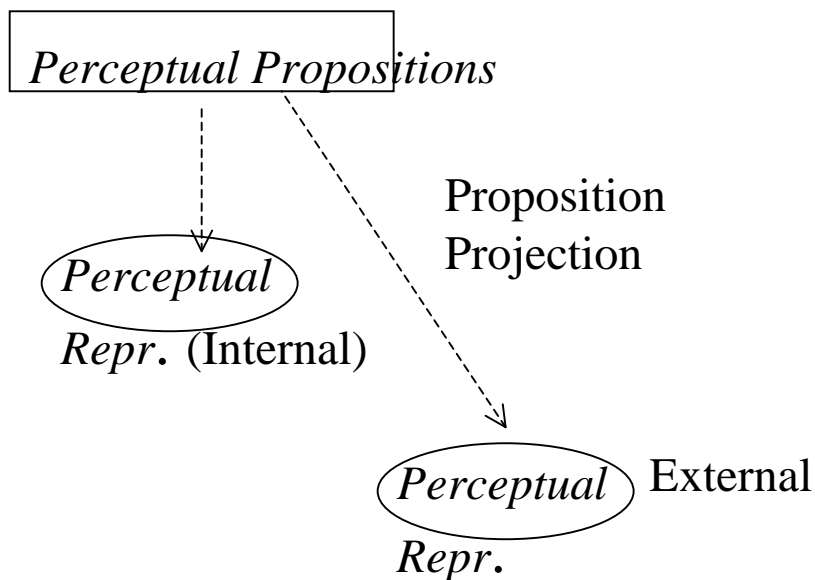
Proposition Extraction/Projection

Perceptual Propositions

Perceptual Repr. (Internal)

Proposition Extraction

Perceptual Repr. (External)
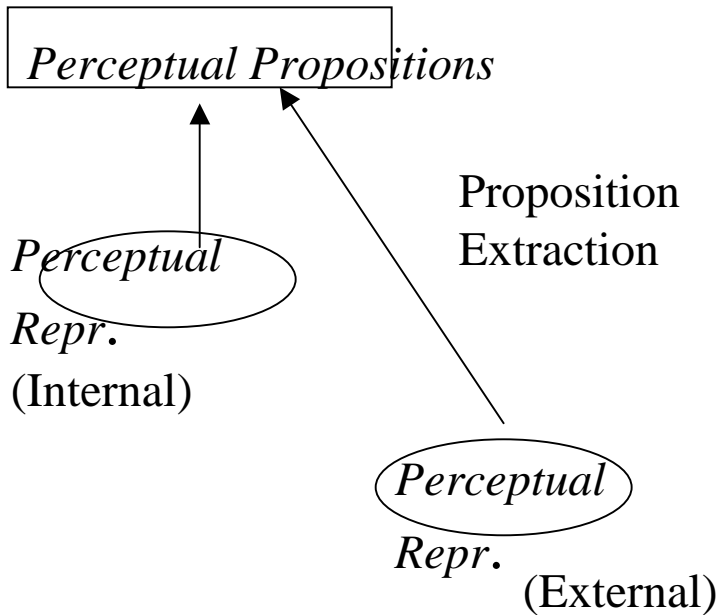
N-ary perceptual propositions

unary: There is an object, object is umbrella, umbrella is soft, mouse is hard to hold

Binary: Umbrella left-of table

Perceptual Propositions

Perceptual Repr. (Internal)

Proposition Projection

Perceptual Repr. External
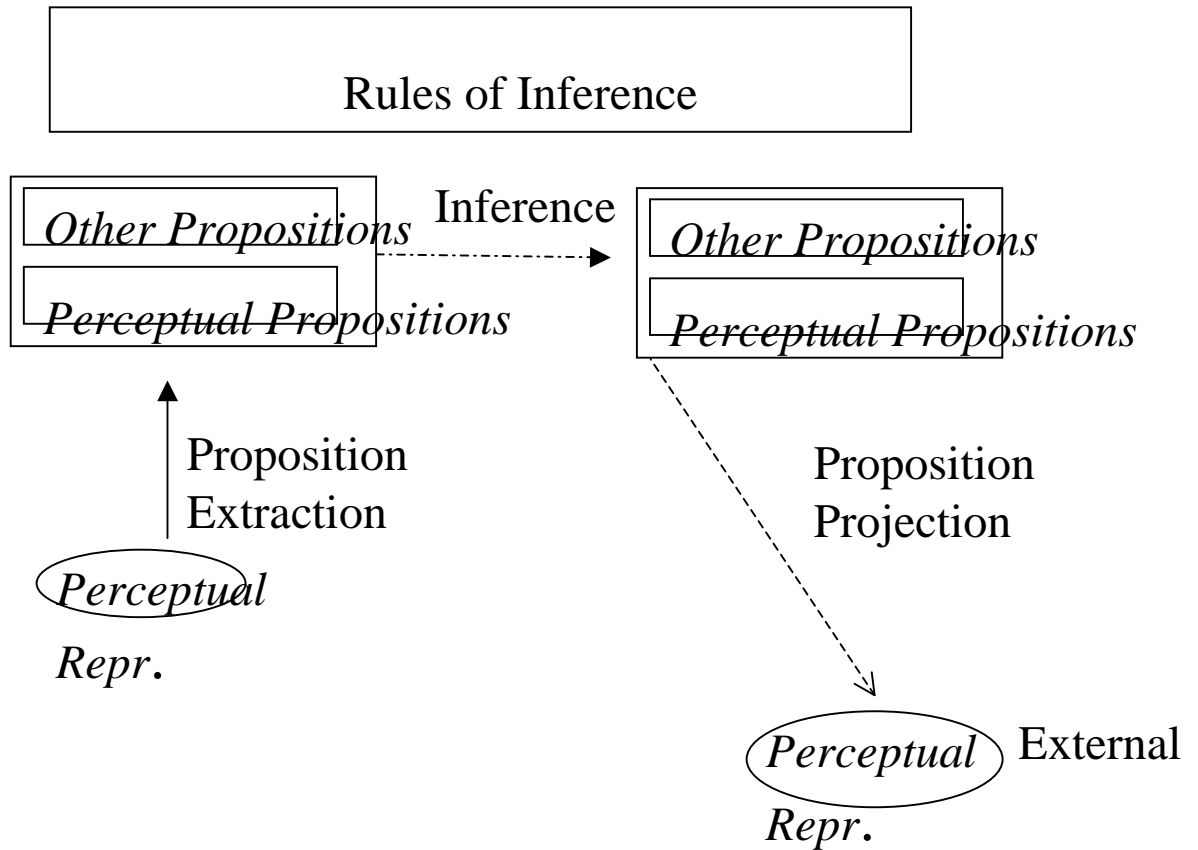
Simulation

- External representations can often be moved, rotated, etc., And propositions about the resulting state can be extracted.
  - *Saves a whole chain of inferences.*
  - This is often the source of the great potency of "using the world as the representation."

Reasoning

Rules of Inference

Other Propositions   Inference   *Other Propositions*

Perceptual Propositions   *Perceptual Propositions*

Proposition
Extraction

Perceptual
Repr.

Proposition
Projection

Perceptual   External
Repr.

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 predicates available for extraction

Reasoning (Inference-Making) (contd)

– Propositions extracted from perc. Reps often replace computationally more expensive inference chains.

- Defining the conditions under which the the propositions that are extracted correspond to correcet 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.

The "Standard" View: Cognition is a "Symbolic" System

- The "standard" or "canonical" view of cognition in traditional CogSci/AI is that it is a *symbol-processing* system.

  - What exactly does this mean?

- This notion of symbol system is inspired by two sets of ideas:

    - NL, or something close to that, is the medium of "thought." (Language of thought hypothesis.)

      - From "thoughts have propositional content," to "thoughts are representations of propositions."

      - (Propositions are assertions about some world, real or imagined, concrete or abstract)
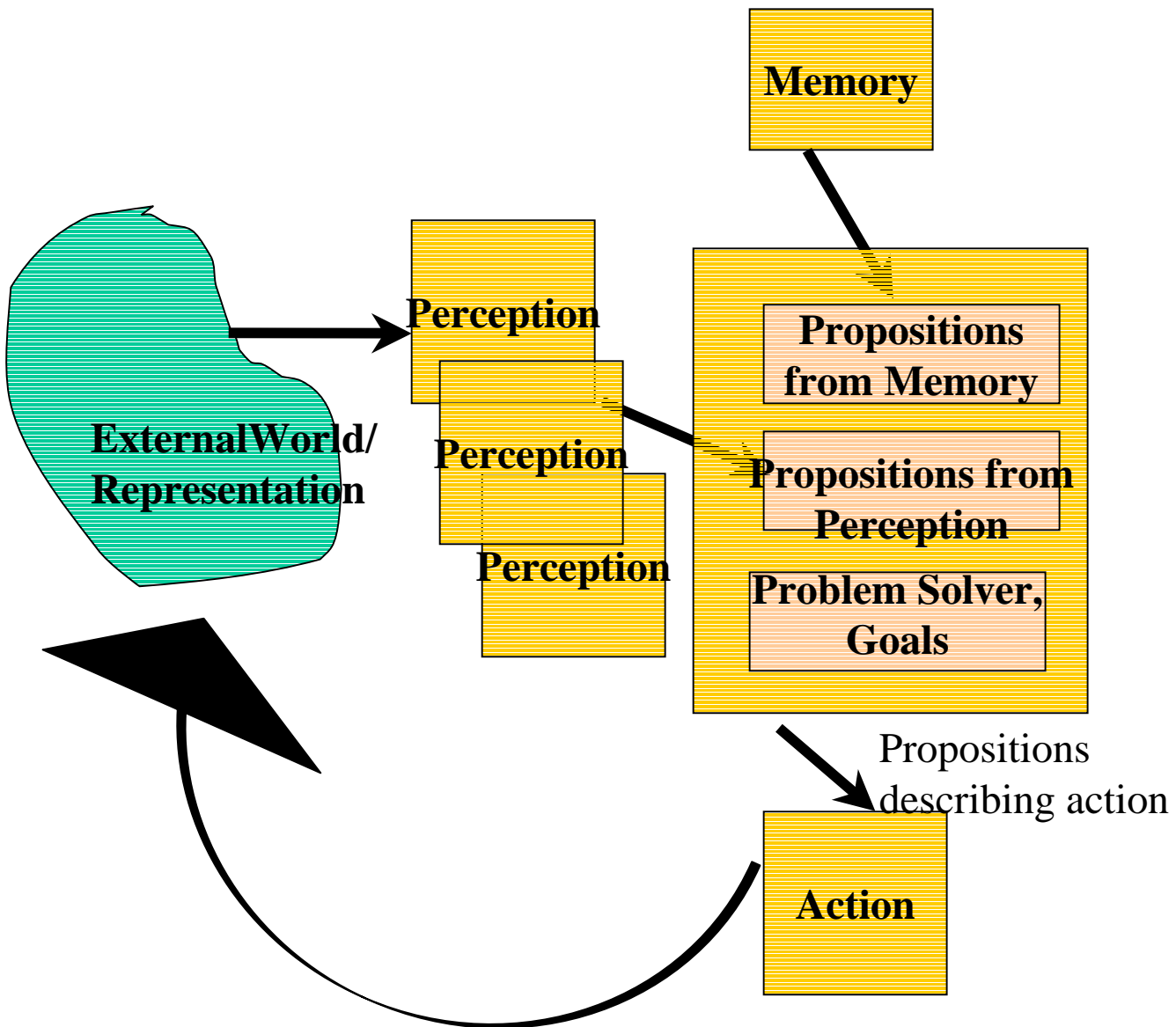
The "Standard" View -continued

- The second inspiration was from the idea of Turing machines and modern computers
  - "Symbols" here are a specific type, "Turing symbols"
  - Turing symbols have the property that the processing of them is purely "syntactic."
- Thus was born the extraordinarily productive notion of thinking as not simply information processing, not simply representation-processing, but more specifically "Turing symbol processing".
- 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."
- AI and cog Sci view cognition as a phenomenon involving concepts.

The "Standard" View and Interaction with the
External World

- 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.

The Standard Story in AI

Memory

Perception

Perception

Perception

ExternalWorld/
Representation

Propositions
from Memory

Propositions from
Perception

Problem Solver,
Goals

Propositions
describing action

Action

What is really in a Cognitive State?

- 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 out 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

Memory

Perception

ExternalWorld/ Representation

IPS (Image)
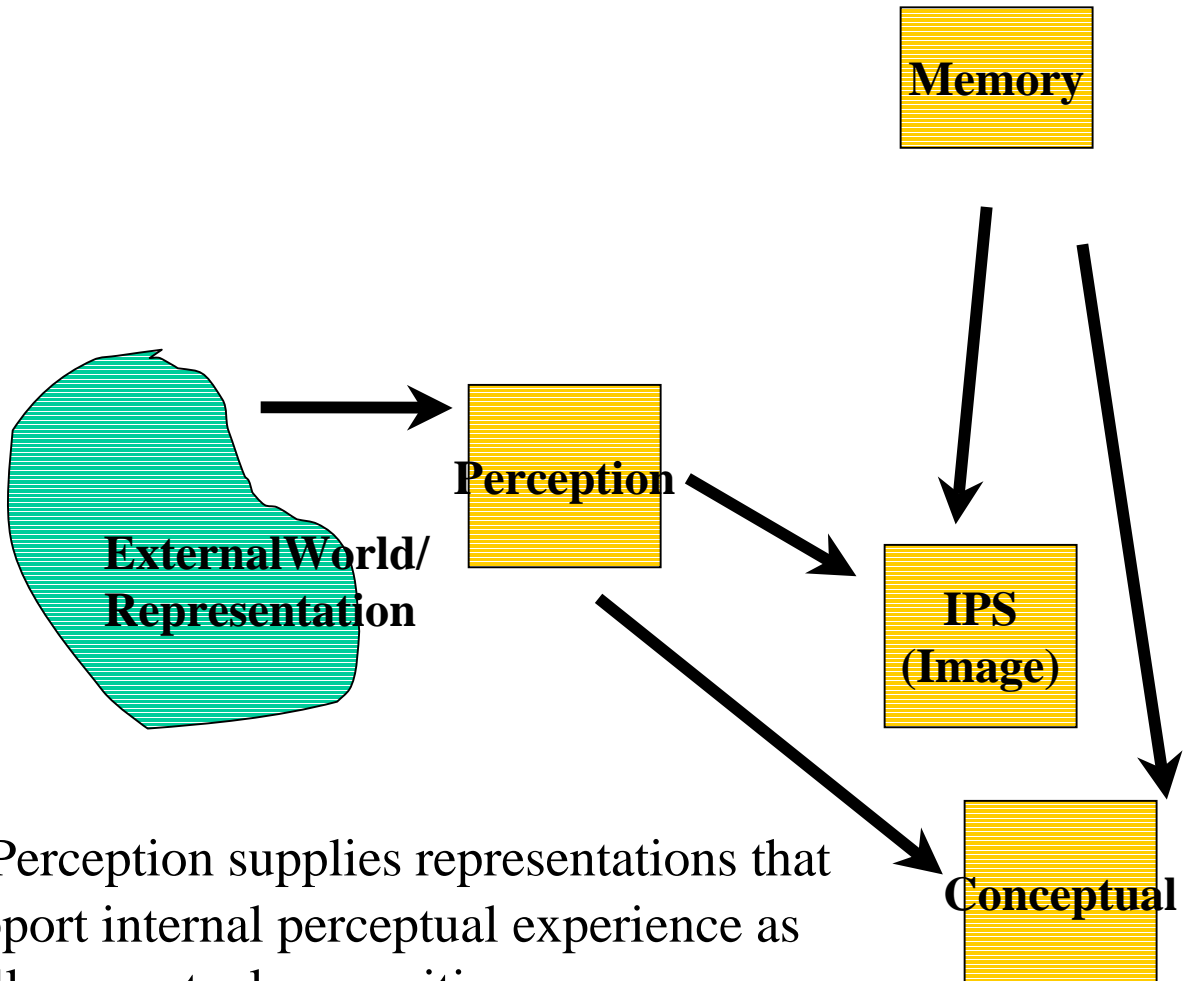
Conceptual

-- 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."

Cognitive State is Multimodal

- Each state of thinking is *potentially* multimodal.
  - Potentially, because not all instances have all modalities.
- I'll first propose an architecture along with an account of how thinking in this framework is accomplished.
- I'll also describe what this might mean to AI, including approaches to and open issues in implementing such a system.

External Vs Internal Representations

- Internal
  - Representations that are posited to explain mental images
    - Not restricted to the visual modality
    - One can imagine moving through the world, songs being played, objects being manipulated
  - Controversy is not over the phenomena themselves, but over the mechanisms posited to explain the phenomena
    - I propose to take a **functional** view, I.E., Identify the *capabilities associated with the image phenomena*, but avoid deep commitments to mechanisms.

Internal Multimodal Representations:
A Proposed Functional Architecture

- Perceptual modalities (PM) includes kinesthetic modality for this discussion

More than One Perceptual Modality
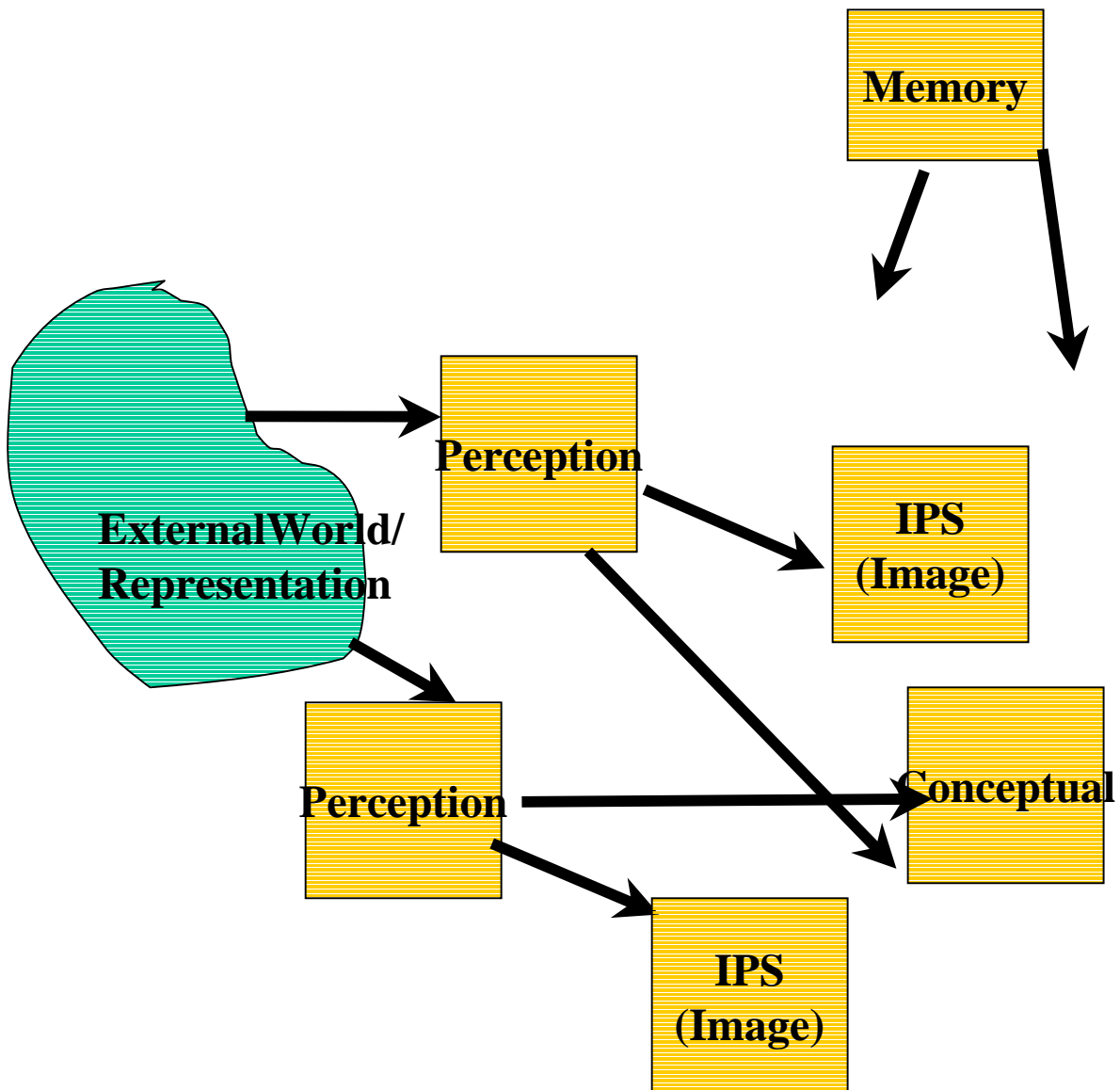
-- Each modality supplies both an image that is experienced as well conceptual predicates

-- 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.

Memory is Multimodal Too

Memory

IPS1

IPS2

IPSn

Conceptual
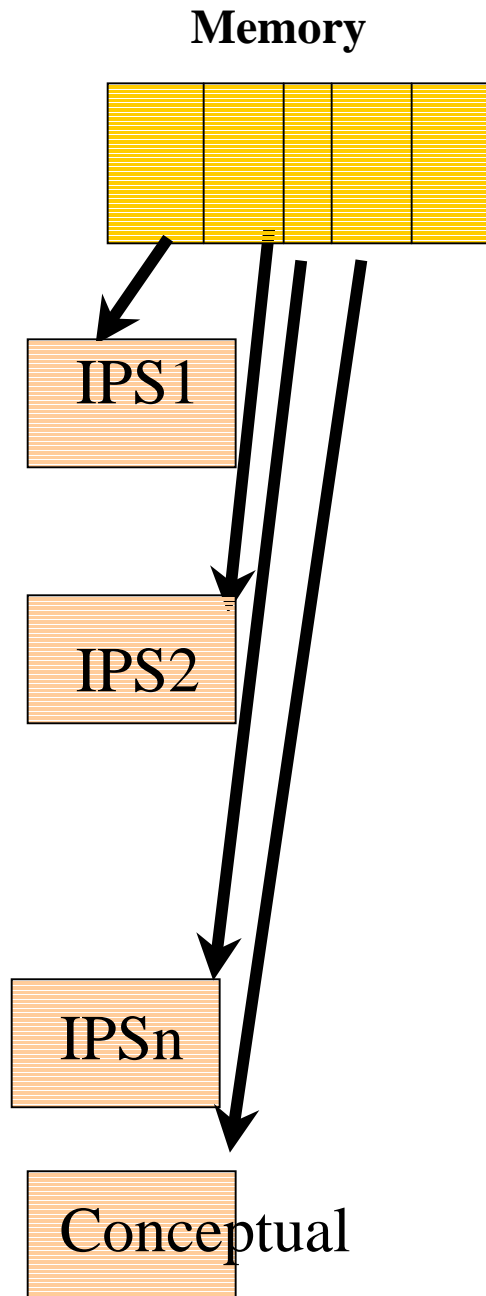
Functional Architecture: Memory is multimodal too

- 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

Intermodal Evocations

Concepts about heart

Auditory images of heartbeats
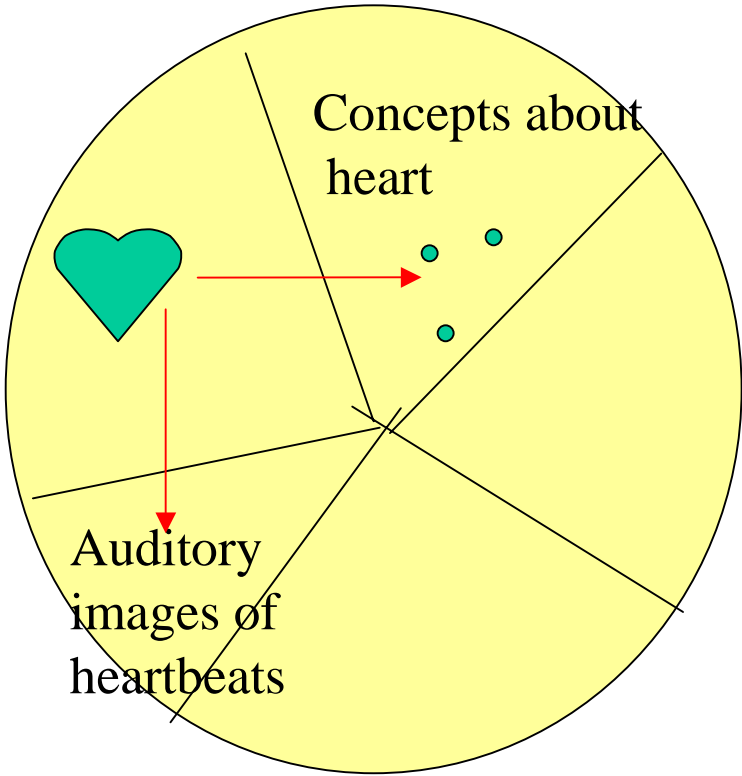
# Representations May Evoke Associated Representations

- 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

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

Memory

*

*

R IPS1

R2 IPS2

IPSn **Rn**

Rc

Conceptual
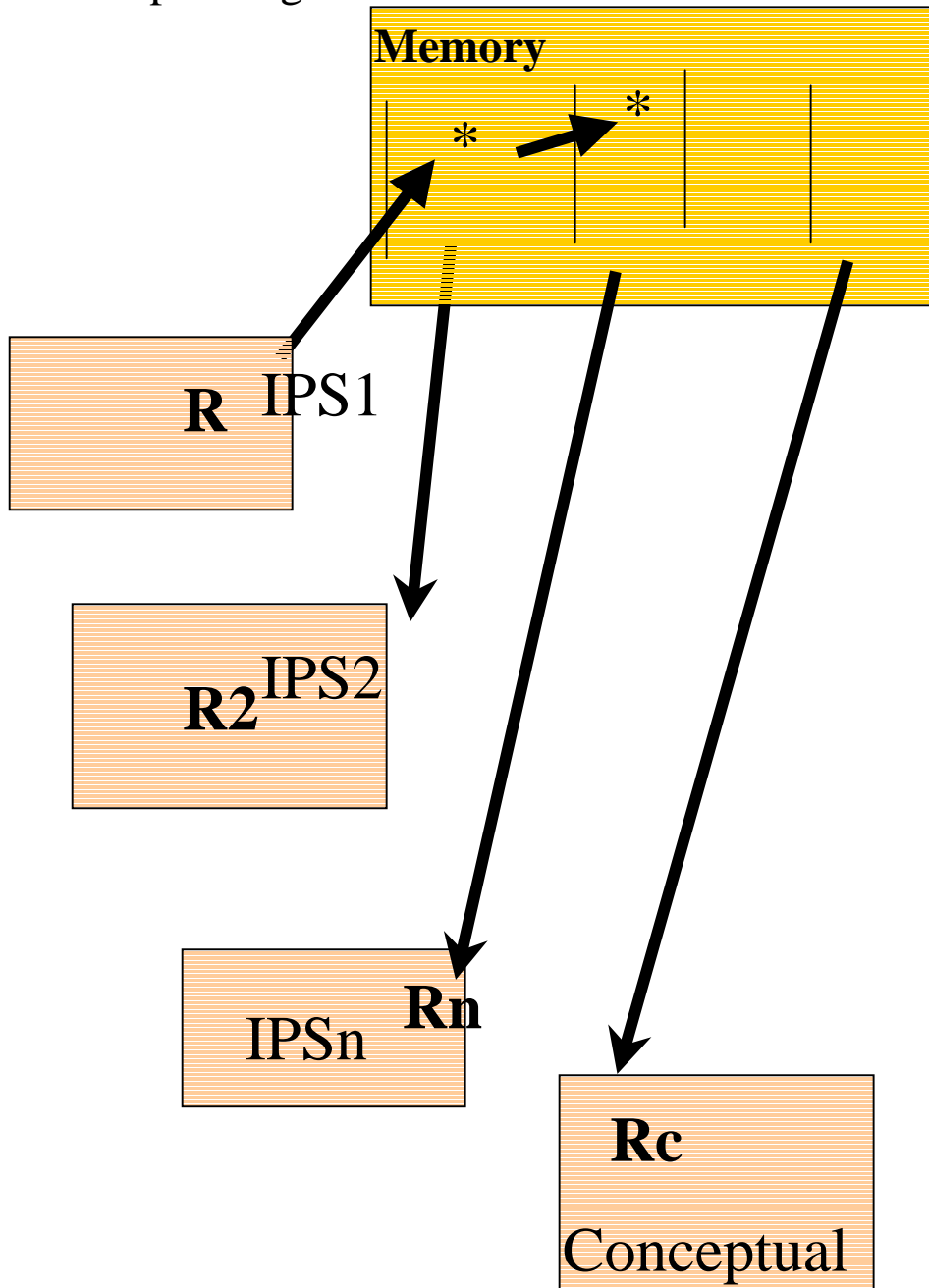
Cognitive state of agent is multimodal
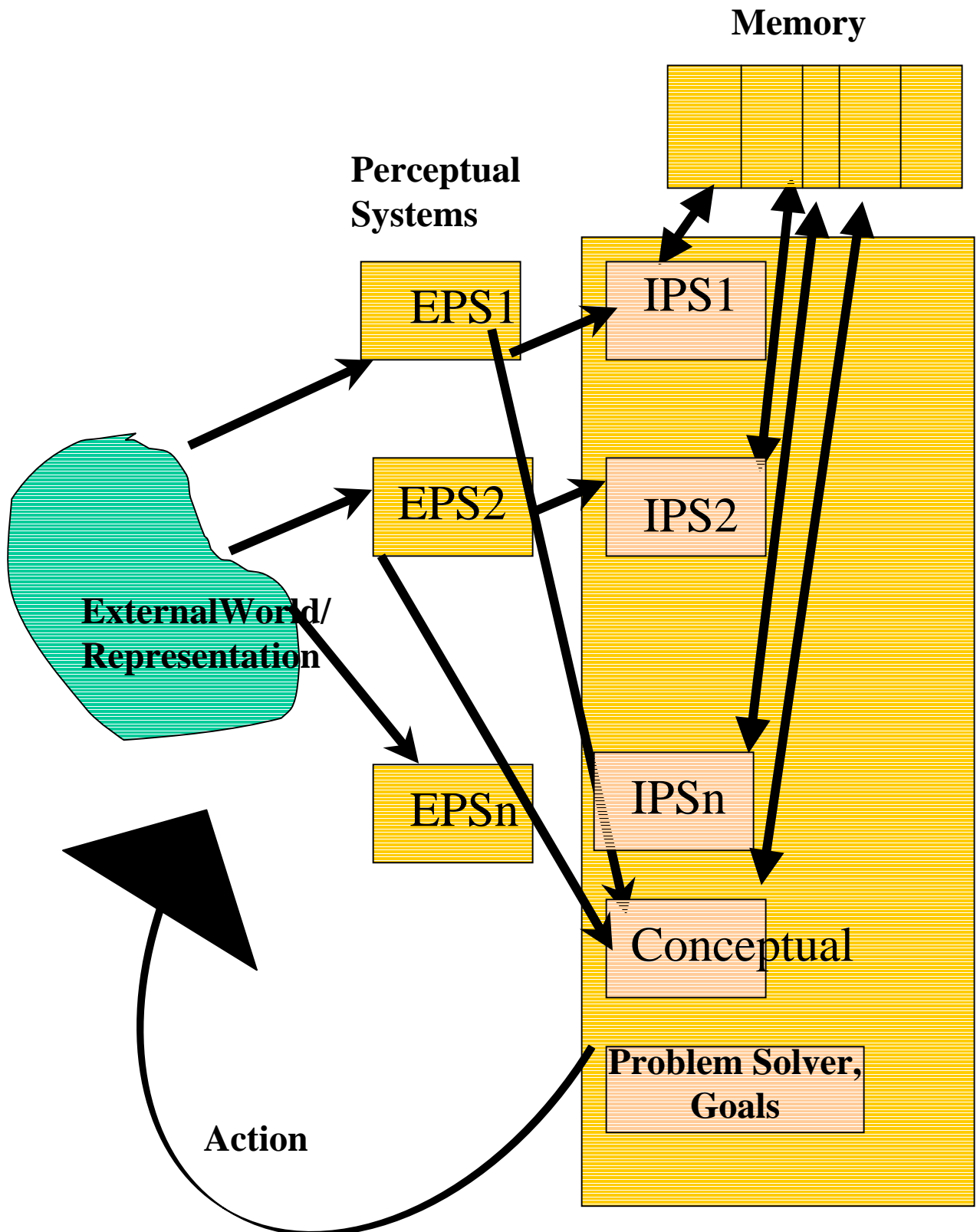
The Architecture is a Functional One

- 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 propostion extractions, while the conceptual modality supports "reasoning"

    - 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.

Issues in the Imlementation/Realization of IPS's

- Still, the basic questions remain:
  - How can IPS's exist in any form other than propositional and symbolic? Clearly, what is in memory cannot be images, since if they are, who will perceive them?
  - If they are images, how can they be composed in novel ways, such as imagining "an elephant eating an apple."

A Brief Outline of the Solution

- Marr's theory of perception
- Image primitives
  - Open theoretical issue

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.